

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

Art Unit: 1645

Priority

Please Note: The elected SEQ ID Nos 1, 3 and 5 do not evidence 100% sequence identity with the sequences set forth in the parent application 08/123, 975. Original descriptive support in the parent application, 08/123, 975 was not found for SEQ ID Nos 1, 3 and 5, nor the elected invention of the combination of SEQ ID Nos 1, 3 and 5. The instantly claimed invention is therefore afforded the priority date of May 12, 1999 (provision application) which evidences original descriptive support for each of these sequences.

THIS PAGE BLANK (USPTO)

ORGANISM Clostridium botulinum
Bacteria: Firmicutes: Bacillus/Clostridium group; Clostridiaceae:
Clostridium.
REFERENCE 1 (bases 1 to 4835)
AUTHORS Bin, T., Kurazono, H., Wille, M., Frevert, J., Wernars, K. and
Niemann, H.
TITLE The complete sequence of the botulinum type A neurotoxin and its
comparison with other Clostridial neurotoxins
JOURNAL J. Biol. Chem. 265, 9153-9158 (1990)
MEDLINE 90264400
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by H.Niemann, 29-NOV-1989.
FEATURES
Location/Qualifiers
source 1..4835
/organism="Clostridium botulinum"
/db_xref="taxon:1491"
240..>4835
/product="neurotoxin mRNA"
344..349
/note="ribosome binding site"
358..4248
/note="neurotoxin"
/codon_start=1
/transl_table=11
/protein_id="AAA23262.1"
/translation="MPYVKQPNKQDPVNGVDIAIKIPNAGOPVKAFAKHNKIMV
IPERDTNPEEGDLPPPEAKQVPVSYDSTYSITDNEKDYLLGVTKLEPRTSTD
LGRMLTIVIRGIPFMGSGTIDTELKVIDTNCINVTQDGSYRSEELNVIIGPADID
IOFECKSFGEHLNLTJRMNGYSTGYIRFSDPTFSELEVDIAPLQAGKAFADPA
VTLAHELHAGHRLGIAINPVRKVMNAYEMSGSEVSEELRTGGGHDARFIDS
LOENFERLYYNNKPKDIASTLNKAKSIYGTASLOXNNPEKELISDPSGKRSVD
KLEPKLTKMLTEITEDNPKFVYLNKRTILANDKAVKINIVPKANNITDYDFNL
NNTNLAANNGONTINNNKFNKLNFTGLFEYLLCVRGITTSKTSLLDGYKAL
NDLCIKVNMWDLFSPSEDNFTNDLNKGEITSDNINAEENISLDLQOYYLFNF
DENENISIEMLSDIIGLELMPNIEFEPNGKYELEKTYWHLRQEFHGKRSRI
ALTNSVNALLNPSRVYTFSSDYKVKAKKAEAMPLGMYQLVDFDESEVST
DRIADITIIIPYIPALINGNMLKDDPGVALIFSGAVILLEFIEIDILPIVGTALV
SYIAKKVLTVOITNALS RNKEMDEVKYYITNMLAVNTOIDILRRKMKALNQA
EATKAIINVOITNTEBEKNINENIDSSKLSNINKAMINIKELKQCSVSLKM
SMIPYGVKRLDEDFDASLDALIKYIDNKGTLIGVDRLKDVNNTLSDIPOLSKY
VDNRLSTFTEYIKNIINISILNLYESNHLIDISRYASKINISKNFIDIDNOI
QLFNLESSKIEVILKNALIVNSMNFSEFWRIRPKYFNSISLNNEVYIICMNNNS
GKWSLNGEIIWTLQDOEIKORVAFKYSOMINISINRIEPTIITNENRNLKXY
INGRLIDCKPISLNGIHASNNIMFKDGCRTDHRVYIKYNNLIDKLENEEIDLY
DNQNSGILKDFWGYLOYDKRYIYMLNLTDPKYYDVANNVGRGYMYLKGPGSVYTT
NLYLNSSLRGTKFIKKYASGNKDNINRNDRVINVVAKKEYRLATNMQAQAEVK
ILSALEIPDVGNLSQVYVMSKNDGIVTKCKMNLQDNNNDIGEIFGHQFNINAKLV
ASNNYROIERSRILGCSWEFIIPVDGGERPL"

BINZ 240
1990

misc_feature /note="potential terminator: putative"
BASE COUNT 1934 a 517 c 756 g 1628 t
ORIGIN

Query Match 55.0%; Score 727.6; DB 1; Length 4835;
Best Local Similarity 72.2%; Pred. No. 8e-177;
Matches 946; Conservative 0; Mismatches 364; Indels 0; Gaps 0;

QY 14 tgcctacccctgaatgatacacaagaacatcacaacccctgaacccgctgcgc 73
DB 2945 TATCTCACTTTACTGAAATATTAAGAAATATTATTAATCTTATTAATGAATTTAAGAT 3004
QY 74 acgaatcaatcaacacgacgacccgctgcgcacgctcccaaatcaaatcggtctca 133
DB 3005 ATGAAGTATATCATTTAATAGACTTATCTAGATGATGCATCAAAATAAATATGTAGTA 3064
QY 134 aagttacatcgatccgacgacgaagaatcaatcagcgtgtcaatcgtgaactctcca 193
DB 3065 AAGTAAATTTTGGATCCAAATAGATAAAATCAATTAATTAATTAAGAAAGTAA 3124
QY 194 aatcgaaatcaccctgaagaatgcatcgtatcaaacctcgtcgaataactctcca 253

DB 3125 AAATGAGTAATTTTAAAAAATGCTATTGTATATATAGATGATGAATAATTTTAGTA 3184
QY 254 cctccttcggtacgcatcccggaatcactcaaccctcctcgtgaacaatgataca 313
DB 3185 CTAGCTTTTGGATTAAGAAATTCCTTAAGTATTTAAACGTATTAAGTCAATTAATGAATA 3244
QY 314 ccatcaacaacgcatggaagaacaactcgtgtggaagatcatcgtgaactcgtgga 373
DB 3245 CAATTAATAATTTGTATGAAATAATTAATTCAGATGAAAGTATCATTAAATTTATGCGAAA 3304
QY 374 tcatcgtgactcgtcgaagaacactcagaagaatcaacaacgctgtgtatcaatactcc 433
DB 3305 TATCTGCACTTTACAGAGTACTCAGGAATAATAACAAGAGTAGTTTAAATACAGTC 3364
QY 434 agatgatacaacatcctgactacatacaatcgcgtgactcgtgtacatcaacaacac 493
DB 3365 AATGATTAATATATATGATTAATTAATAAGAGAGATTTTGTAACTATCACTAAATAA 3424
QY 494 gctcgtataactcccaaatcactacatacaacgctgactgtgacgcaagaacgactcca 553
DB 3425 GATTAAATTAATCTTAATAATTTATATTAATGAAGATTAAATAGTCAAAAACCAATTCCA 3484
QY 554 atcgtgtaacatccacgctctcctaatacatcagttcacaactgacggtgtcgtgaca 613
DB 3485 ATTAGTATATATTTATCATGCTAGTAAATATATATATGTTAAATTAGATGTTGATGAGATA 3544
QY 614 ctacacgcatcacatcgtgatacaatcactcaatcgtgtcgtgacaagaacgcaaga 673
DB 3545 CAATGATGATATATTTTGGATTAATAATTTAAATCTTTTGAAGAAATTAAGAAAG 3604
QY 674 aatcaagaacgctgtcgcgaacacgactccaactcgtgtatccgcaagaagctctg99g 733
DB 3605 AAATCAAGATTTATATGATTAATCAATCAATTCAGGTATTTAAAGACTTTTGGGGTG 3664
QY 734 actaccgtgagtaacgaacaaactactacatcgtcgaatcgtgacgtcgaacaaatac 793
DB 3665 ATTATTTACAAATGATGATTAACCATCTACTATATGTTAAATTATATGATCCAAATAATAT 3724
QY 794 ttgacgtcaacaatcgtgactacccggtgtacatcgtgacgtgaaaggtccggtgtcgt 853
DB 3725 TCATATTAATTAATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTA 3784
QY 854 tctgactacacaacatcactgactcctcctcgtgacggtgtgacaaatcactac 913
DB 3785 TATGACTCAAAATTTATTTAAATTTCAAGTTGATATAGGGGACAAATTTATATATA 3844
QY 914 agaaatacgcgtcgtgtaacaagaagaatacgttcgcaacaatgactcgtgtatcata 973
DB 3845 AAAAATATGCTCTCGAAATAAAGATAATATATGTTAGAAATATATGATCGTATATATTA 3904
QY 974 atgtgtagtaagaacaagaatcaccgtgtcgtacacatgcttctcagcgtgtgtg 1033
DB 3905 ATGTAGTATGTTAAATAAATAAAGAAATAGGTTACTCTAATGCATCAACAGGCGGTG 3964
QY 1034 aaaaagctcgtcgtcgtgtaaatcccggaagctgtgtaactcgtcgtcgtgtgtgaa 1093
DB 3965 AAAAATATGCTAAGTATTAAGAAATACCTGATATAGGAAATTCAGTCAAGTGTGTA 4024
QY 1094 tgaatccaaagaacgcaacggtgatacactcaacaatgcaaaatgaaatcgtcgaagaaca 1153
DB 4025 TGAAGTCAAAAATATGATCAAGAAATTAACAATAATTAATGCAAAATTTATCAAGATATA 4084
QY 1154 atgtgaacatcgtgttccatcgtgtccacacagttcacaacatcgttaaatcgtgtg 1213
DB 4085 ATGGAGATGATATAGGCTTATATAGATTTATCATGATTTAATATAGCAATCAATCTAGTA 4144
QY 1214 ctccaactggtacatcgtcgaatcgaatcgtcctcctcgtcactcgtgtgtcgtgtg 1273
DB 4145 CAAGTATTTGCTTAATATAGCAAAATAGAAAGATCTGATGAGACTTTGGGTTGCTCATGG 4204
QY 1274 agttcaccggtgtgactgacggtgtgtggaagcgtcgtcgtgaagaatc 1323
DB 4205 AATTTATCTGTATGATGATGATGGGAGAAAGCCACTGTATTAATATC 4254

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

Db 432 TCAGATGATCAACATCTCTGCTACATCAATTCGCTGATCTTGCTTACCAATCAACAA 491
 Qy 492 tcgctgaataactccaataatctacatcaacggcgctctgatcgaccagaacgatctc 551
 Db 492 TCGCTGAATTAACATCAAAATCTTACATCAACGGCGCTCTGATCGACCAAGAACGATCTC 551
 Qy 552 caatctgggtgaatacaccagctcttaataacatcatgttccaactggagctgtgtcgtga 611
 Db 552 CAATCTGGGTGAATCACTCAGCTCTTAATTAACATCATGTCTCAAACTGGACGGTGTCTCTGA 611
 Qy 612 cactcaccgctacatctcggttcaaatctcaatctgttctgcgaagaagaactggacgaaa 671
 Db 612 CACTCACCCTGATCATCTGGATCAAAATCTTCAATCTGTGCAAAAGAACTGGAAGCAAAA 671
 Qy 672 agaatcaagaagctgtgacgaacacgctccaatctctgtatctcttgaagaactctggg 731
 Db 672 AGAATCAAGAGCTGTGACACAAACCACTCAATCTTGTGATCTCTGAAAGACTTCTGGGG 731
 Qy 732 tgaactactgacgtacgacaaacgctactacatgtctgtaactgtgacatccgaacaata 791
 Db 732 TGACTACTGACGTACGACAAACCGCTACTACATGCTGAATCTGTACGATCGAACAATA 791
 Qy 792 cgttgacgtcaacaatgtatcggttaccggtttacatgttaccatgaagctcgggttc 851
 Db 792 CGTTGACGTCAACAAATGTAGGTATCCCGGCTTACATGTACCTGAAAGCTCGCGTGGTTC 851
 Qy 852 tttatgacacccaacatctaccctgaactctccctgtaccggtgtacgaataatcatcat 911
 Db 852 TGTATGACTACCAACATCTTACTGAACTCTTCCCTGTACCGTGTGACCAATTCATCAT 911
 Qy 912 caagaataacgctctgtgtacaaagaacatctctgtcgaacaatgatcgtgtatacat 971
 Db 912 CAAGAATACGCGCTGTGTAACAAGAACAAATATGTTGCGAACAAATGATGTATATCAT 971
 Qy 972 caatgtgtgttaagaacaagaataacgctctgtgtaccatgaatctcgaagctgtgtc 1031
 Db 972 CAATGTGTGTAAAGAACAAAGAAATACCGTCTGCTACCAATGCTTCTCAGCGTGGTGT 1031
 Qy 1032 agaaaagaatctgtctctctgtgaatacccggaagctgtgaatctgtctgaagtaagt 1091
 Db 1032 AGAAAAGATCTTGTCTCTCTGGAATCCCGGAGGTGTGTAATGTCTCAGGATGTTGT 1091
 Qy 1092 aatgaataacgaagacagcaggttaccatcaacaaatgcaaaatgtacatctgcagagcaa 1151
 Db 1092 AATGAATACGAAGACAGCAGGTATCACTTAACAATGCAAAATGATCTGCGAGCAAA 1151
 Qy 1152 caatgtgaacgatacgggttccatcggcttccaccagctcaacaataatcgctaaactgt 1211
 Db 1152 CAATGTGAACGATATCGGTTTCATCGGTTTCCACAGTTCAACAATATCGCTAAACTG 1211
 Qy 1212 tgcctgaactgtgacaaatctgcagatcgaaactctctcgcagactctgtgtgtcttg 1271
 Db 1212 TGCTTCCAACTGTGACAAATGCTGATGAGACGTTCTCTGCGACCTGTGGGTTCTCTTG 1271
 Qy 1272 ggaatcattcccggttgaatgacggttgggggtgaacgtcgcgtgtga 1317
 Db 1272 GGAGTTATCCCGGTTGATGACGGTTGGGGTGAACGTCCCGCTGTA 1317
 RESULT 3
 AX036243 1330 bp DNA linear PAT 16-NOV-2000
 LOCUS AX036243 Sequence 22 from Patent EP041149.
 DEFINITION AX036243
 ACCESSION AX036243
 VERSION AX036243.1 GI:11225862
 KEYWORDS
 ORGANISM
 SOURCE
 ORGANISM
 REFERENCE
 AUTHORS Kink,J.A., Flrica,J.R., Padhye,N.V., Thalleay,B.S., Stafford,D.C. and
 WILLIAMS,J.A.
 TITLE Vaccine and antitoxin for treatment and prevention of C. Difficile

JOURNAL disease
 Patent: EP 1041149-A 22 OCT-2000;
 OPIDIAN PHARM INC (US)
 FEATURES Location/Qualifiers
 source
 1.1330
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 1.1317
 /note="Synthetic"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC16478.1"
 /db_xref="GI:11225863"
 /translation="MARLSTPEYIINKIINTSILNRYESNHLIDLSRYASKINIGS
 KVEDPDKNOIQLFNIESKIEVILKNAIVNSYENFSTFPIRIKPFNSILN
 EYTLINMENSNGKXVSLNGEITLMDLOEIKOYHVFYSOMINSDYINRIVFT
 ITNRLNLSKIYINGRLIDKPSISNLSNTHASNNIFKIDGCDTHRYIWKIFNLPD
 KELNEKEIKLDYDNOSSGILKDPFWDYDLYDKPYTMDLNDYDNKYDVANNGVIGY
 YLGRPSVMTNIVLNSLYRGTKEPIKKYVAGNDNIYRNDRYIVNVVNNKRYR
 LATNASQAGEKILSLAIEIPVGNLSQVVMKSKNDGITNKCKMLDNDNGNDIGFI
 GFHGFNNIAKLIVASNMVNRQIERSSRTLGCSWEPFIPVDGMBRPL"
 BASE COUNT 400 a 339 c 246 g 345 t
 ORIGIN
 Query Match 98.6%; Score 1304.4; DB 6; Length 1330;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 12 gatgtcactctcactgaataacatcaagaacatcatcaatcctccatcctgaacctgcg 71
 Db 12 GCTGTCACTTCACTGATGATCAATCAAGAACATCATCATCAATCTGATCTGACGCG 71
 Qy 72 ctacgaatccaatctactgtatcgacgtgtctcgtctgaagcttccaaataacatcggttc 131
 Db 72 CFACGAATCCAAATCACTGATCGACCTGTCTGCTGACGCTTCCAAATCAACATCGGTT 131
 Qy 132 taaagtttaactctgacatcgacacgaagaacatcagatcagctgttcaatctgaaatcttc 191
 Db 132 TAAAGTTTAACTTCATCGATCGATCGACCAAGAAATCAATCAAGCTGTTCATCTGGAATCTTC 191
 Qy 192 caaatcgaagttaacctgaagaatgtctatcgtatacaactctatgtacgaaaactcttc 251
 Db 192 CAATATGAATTAATTCGGAAGAAATGATGTATACAACTGTATGTAAGAAATCTTCTC 251
 Qy 252 caactccttctgtatcgtatccggaatctcaactcaactcctcctcgtgaacatgaata 311
 Db 252 CACTCTCTTCTGATCCGTATCCCGAAATACTTCAACTCACTCTCTGGAACAATGAATA 311
 Qy 312 caccatcaatcaactgtcgtggaagaacatctgtgttgaagaatctcgtgaactcgttga 371
 Db 312 CACCATCAATCAACTGATGATGGAAGAAATCTGTGTTGGAAGATTAATCTGAAACTCGGGGA 371
 Qy 372 aatcatctgacatctgcagagacactcaggaatcaacaacaggtgttgtatcaaatcctc 431
 Db 372 AATCATCTGACCTCTGAGGACACTCAAGGAATCAACACGCTGTGTATCAATATCTC 431
 Qy 432 tcaagatgatacaaatctctgactacatcaatcgtctgtatctcgtttaccatcacaaca 491
 Db 432 TCAGATGATCAAACTCTCTGACTACATCAATCGCTGTGATCTCTGTTACCAACACCAAA 491
 Qy 492 tcgctgaataactccaataatctacatcaacggcgctctgatcgaccagaacagatctc 551
 Db 492 TCGTCTGAATTAACATCAAAATCTTACATCAACGGCGCTCTGATCGACCAAGAACGATCTC 551
 Qy 552 caatctgggtgaatacaccagctcttaataacatcatgttccaactggagctgtgtcgtga 611
 Db 552 CAATCTGGGTGAATCACTCAGCTCTTAATTAACATCATGTCTCAAACTGGACGGTGTCTCTGA 611
 Qy 612 cactcaccgctacatctcggttcaaatctcaatctgttctgcgaagaagaactggacgaaa 671
 Db 612 CACTCACCCTGATCATCTGGATCAAAATCTTCAATCTGTGCAAAAGAACTGGAAGCAAAA 671

THIS PAGE BLANK (USPTO)

VERSION AM06241.1 GI:11225862

KEYWORDS

SOURCE synthetic construct.

synthetic construct.

artificial sequence.

REFERENCE 1 (bases 1 to 1330)

AUTHORS Kink, J. A., Pitra, J. R., Padhye, N. V., Thalley, B. S., Stafford, D. C. and

Williams, J. A.

TITLE Vaccines and antitoxin for treatment and prevention of *C. difficile*

disease

JOURNAL Patent: EP 1041149-A 22 Oct-2000;

OPHIDIAN PHARM INC (US)

FEATURES Location/Qualifiers

SOURCE

1..1330

/organism="synthetic construct"

/db_xref="taxon:32630"

CDS

1..1317

/note="Synthetic"

/codon_start=1

/transl_table=11

/protein_id="CAC16478.1"

/db_xref="GI:11225863"

/translation="MARLLSTPTEYIKNIINTSILNLRYESNHLIDLSHYASKINIGS
RVNFDPIKNOIQLFNLESSKIEVILKNAIVYNSMYENFTSFWRIPKYFNSISLNN
EYTIINCMEHNSGKWSIHYGEIILWLODQTEIKQVYVFKYSQMINISQYINHWIFPT
ITNRLNWSKTYIGRLIDQKPIISLQKIHASNNIMFKLDDCQDTHRYIWKYFNLFD
KELNKEIKDLVINDOSNSGILKHPWGVYQDKPYMYLNDYINPKYVIVNVGIRGVM
YLKGPVGSVMYTTNLYLNSLYKTKF1KKYASGNKUNIVKNDORVYINVVKNKKEYR
LATNASOAGVEKILSALEIPDVGNLSOVVVMKSKNDGQITNKKCMNLDQNGNDIGFI
GFHQFNNAIKLVASWYNRQIERSSRTIGCSWEFIPVDGWERPL"

BASE COUNT 400 a 339 c 246 g 345 t

ORIGIN

Query Match 98.2%; Score 1302.2; DB 6; Length 1330;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Oy 14 tggcctcctacccttcaactgaatcacatcaagaacatcatcaatcctccatcctgaacctgc 73
Db 11 TGCTGTCTACCTTCACTGAATACATCAAGAACATCATCAATACCTCCATCCTGAACCTGC 70

Oy 74 gctacgaatccaatcacctgatcgaccctgctcgtctacgcttccaaaatcaacatcggtt 133
Db 71 GCTACGAATCCAATCACCTGATCGACCTGTCTCGCTACGCTTCCAAAATCAACATCGGTT 130

Oy 134 ctaaagttaacttcgatccgatcgacaagaatcagatccagctgttcaactctggaatcctt 193
Db 131 CTAAGTTAACTTCGATCCGATCGACAAGAATCAGATCCAGCTGTTCAATCTGGAATCTT 190

Oy 194 ccaaatggaagttaacttcgatccgatcgacaagaatcagatccagctgttcaactctggaatcctt 253
Db 191 CCAAAATCGAAGTTATCTTGAAGATGCTATGTTATACAACCTTATGTACGAAAACCTTCT 250

Oy 254 ccacctccttctggatccgatccggaataacttcaactccatctctctgaacatgaat 313
Db 251 CCACCTCTCTTGGATCCGATCCGGAATACTTCAACTCCATCTCTCTGAACATGAAT 310

Oy 314 acaccatcatcaactgcattggaataacttctggttggaaagtatctctgaactacgggt 373
Db 311 ACACCATCATCAACTGCATGGAAACAATCTTGGTTGGAAAGTATCTCTGAACATCGGTT 370

Oy 374 aaatcatctggactctgcaggacactcaggaatacgaacagcgtgttgattcaataact 433
Db 371 AAATCATCTGGACTCTGCAGGACACTCAGGAATCAAAACGCGTGTGTATTCAAATACT 430

Oy 434 ctacagatgatcaacatctctgactacatcaatcgctggatctctgttaccatcaccaaca 493
Db 431 CTCAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTTCTGTACCATCACCAACA 490

Oy 494 atcgtctgaataactccaaaatctacatcaacggcgtctctgatcgaccagaaacggatct 553
Db 491 ATCGTCTGAATAACTCCAAAATCTACATCAACGGCGCTCTGATCGACAGAAACCGATCT 550

Oy 554 ccaatctgggttaactccacgcttctaataacatcatgttcaactggacggtgtgtcgtg 613
Db 551 CCAATCTGGGTAAACATCCACGCTTCTAATAACATCATGTTCAAATGGACGGTGTGTCTGT 610

Oy 614 acactcaccgctacatctggatcaataacttcaatctgttgcacaaagaactgaacgaaa 673
Db 611 ACACTCACCGCTACATCTGGATCAAAATCTTCAATCTGTTCGACAAAGAACTGAACGAAA 670

Oy 674 aagaaatcaagacctgtacgacaaccagtcgaattctggtatcctgaagacttctggg 733
Db 671 AAGAAATCAAAGACCTGTACGACAACCGTCCAATCTCGTATCTGAAAGACTTCTGGG 730

Oy 734 gtgactacctgcagtcagcaaaacggtactacatgctgaatctgtacgaccgaacaaat 793
Db 731 GTGACTACCTGCAGTACGACAAACCGTACTACATGCTGAATCTGTACGATCCGAACAAAT 790

Oy 794 acgttgacgtcaacaatgtaggatccgggttacatgtacgtgaaggtccgcgtggtt 853
Db 791 ACGTTGACGTCAACAATGTAGGTATCCCGGTTACATGTACCTGAAAGGTCCCGGTGGTT 850

Oy 854 ctgttatgactaccaacatctacgtgaactcttccctgtacogtgatccaatcactca 913
Db 851 CTGTTATGACTACCAACATCTACCTGAACCTTTCCTGTACCGTGGTACCAATTCATCA 910

Oy 914 tcaagaatacgcgtctggttaacaaggacaatctcgttcgcaacaatgatcgtgtatata 973
Db 911 TCAAGAAATACGCGTCTGGTAACAAGGACAATATCGTTCGCAACAATGATCGGTATACA 970

Oy 974 tcaatgttgtatgaagacaaagaatcgttctggctaccaaagcttctcaggtctggt 1033
Db 971 TCAATGTGTAGTTAAGAACAAAGAAATACCGTCTGGCTACCAATGCTTCTCAGGCTGGTG 1030

Oy 1034 tagaaagatcttctctgctctggaataccggacgttggttaactctgtctcaggttagttg 1093
Db 1031 TAGAAAAGATCTTGTCTGCTCTGGAAATCCCGGACGTTGGTAATCTGTCTCAGGTAGTTG 1090

Oy 1094 taatgaatccaagaacgaccagggtatcactaacaatacgaaaatgaatctgcaggaca 1153
Db 1091 TAATGAATCCAAGAACGACCGGATCACTAACAATGCAAAATGAATCTGCAGGACA 1150

Oy 1154 acaatggttaacgatatacgttctcatcggtttccaccagttcaacaatactgcataactgg 1213
Db 1151 ACAATGGTAACGATATCGGTTTCATCGGTTTCCACCAAGTTCAACAATATCGCTAAACTGG 1210

Oy 1214 ttgcttcaactggtacaactcgtcagatcgaaacttctctcgcactctgggttgccttt 1273
Db 1211 TTGCTTCAACTGGTACAATCGTCAGATCGAACGTTCTCTCGCACTCTGGGTTGCTCTT 1270

Oy 1274 gggagttcatcccggttgatgacgggttgggggtgaacgctccgctgtaa 1320
Db 1271 GGGAGTTCACTCCCGGTTGATGACGGTTGGGGTGAACGCTCCGCTGTAA 1317
```

THIS PAGE BLANK (USPTO)

RESULTS
ARI69140 LOCUS AR169140 1330 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 22 from patent US 6290960.
ACCESSION AR169140
VERSION AR169140.1 GI:17906909
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1330)
AUTHORS Kink, J.A., Thalley, B.S. and Stafford, D.C.
TITLE Vaccine and antitoxin for the treatment of C. difficile disease
JOURNAL Patent: US 6290960-A 22 18-SEP-2001:
FEATURES
Location/Qualifiers
source 1..1330
/organism="unknown"
BASE COUNT 400 a 339 c 246 g 345 t
ORIGIN

SEQ ID

Query Match 98.2%; Score 1302.2; DB 6; Length 1330;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 14 tggcctctaccttcactgaatcacatcaagaacatcatcaatccctccatcctgaacctgc 73
Db 11 TGCTGTCTACCTTCACTGAATACATCAAGAATCATCAATACCTCCATCCTGAACCTGC 70

Oy 74 gctacgaatccaatcaccgatcgacctgtctcgctacgcttcccaaaatcaacatcggtt 133
Db 71 GCTACGAATCCAATCACCTGATCGACCTGTCTCGTACGCTTCCAAAATCAACATCGGTT 130

Oy 134 ctaaaagttaacttcgatccgatcgacaagaatcagatccagctgttcaactctggaatctt 193
Db 131 CTAAAGTTAACTTCGATCCGATCGACAAGAATCAGATCCAGCTGTTCAATCTGGAATCTT 190

Oy 194 ccaaaatcgaagttatcctgaagaatgctatcgatatcaactctctgtatcgaaaaactctt 253
Db 191 CCAAAATCGAAGTTATCTCGAAGAATGCTATCGTATACAACCTATGTATGACGAAAACCTCT 250

Oy 254 ccacctctctctggaatccgatccggaatacttcaactccatctctctgaaacatgaat 313
Db 251 CCACCTCTCTCTGGATCCGATCCCGAAATACTTCAACTCCATCTCTGAAACATGAAT 310

Oy 314 acaccatcatcaactgcatggaaaaaacttctggttggaaagtatctctgaactacgggtg 373
Db 311 ACACCATCATCAACTGCATGGAAAAAATCTCGTGGAAAGTATCTGAACTACGGTG 370

Oy 374 aaatcatctggactctcgaggacactcaggaaatcaaacagcggtgttattcaaaact 433
Db 371 AAATCATCTGGACTCTGAGGACACTCAGGAAATCAACAGCGGTGTGTTATTCAAATACT 430

Oy 434 ctccagatgatcaacatctctgaactacatcaatcgctggatctctcgatccatcaacaa 493
Db 431 CTCAGATGATCAACATCTCTGACTACATCAATCGTGGATCTTCGTTACCATCACCAACA 490

Oy 494 atcgtctgaataaactccaaaatctacatcaacggcgctctgatcgacagaaacgatct 553
Db 491 ATCGTCTGAATAAATCCAAAATCTACATCAACGGCGCTCTGATCGACAGAAACCGATCT 550

Oy 554 ccaatctgggttaacatccacgcttctaataacatcatgttcaaaactggacgggtgtcg 613
Db 551 CCAATCTGGGTAAACATCCACGCTTCTAATAACATCATGTTCAAAACGGACGGTGTGCTG 610

Oy 614 acactcaccgctacatctggatcaaaacttcaatctgttcgacaaagaactgaacgaaa 673
Db 611 ACACTCACCCTACATCTGGATCAAAATCTCAATCTGTTCGACAAAGAACTGAACGAAA 670

Oy 674 aagaaatcaagacactgtacgacaaacagtcgaattctggtatcctgaaagactctcg 733
Db 671 AAGAAATCAAGACCTGTACGACAACAGTCCAATCTGGTATCCTGAAAGACTTCTGGG 730

Oy 734 gtgactacctgcagtcacgacaaacgctactacatgctgaatctgtacgacacgaaat 793
Db 731 GTGACTACCTGCAGTACGACAAACCGTACTACATGCTGAATCTGTACGATCGGAACAAAT 790

Oy 794 acgttgacgtcaacatgtaggatcccggttacatgtacctgaaaggtccgcgtggtt 853
Db 791 ACGTTGACGTCAACAAATGTAGGTATCCGCGGTACATGTACCTGAAAGGTCCGCGTGGTT 850

Oy 854 ctgttatgactaccaacatctacctgaactcttccctgtacggtggtacaaattcatca 913
Db 851 CTGTTATGACTACCAACATCTACCTGAACCTTCCCTGTACCGTGGTACCAATTCATCA 910

Oy 914 tcaagaaatcgcgctctggttaacaaggacaatatcggttcgcaacaatgatcggtatata 973
Db 911 TCAAGAAATACGCGTCTGGTAACAAGGACAATATCGTTCGCAACAATGATCGTGTATACA 970

Oy 974 tcaatgtttagttaagaacaaagaataccgctctggctaccaatgcttctcaggctggtg 1033
Db 971 TCAATGTTGTAGTTAAGAACAAGAAATACCGTCTGGCTACCAATGCTTCTCAGGCTGGTG 1030

Oy 1034 tagaaaagatctgtctgctctggaatcccggaactgtgtaactctgctcaggtaggtg 1093
Db 1031 TAGAAAAGATCTGTCTGCTCTGGAATCCCGGACGTGGTAATCTGTCTCAGGTAGTTG 1090

Oy 1094 taatgaaatccaagaacgaccaggtatcactaacaatacgaatgaatctgcaggaca 1153
Db 1091 TAATGAAATCCAAGAACGACAGGATACCTAACAAATGCAAAATGAATCTGCAGGACA 1150

Oy 1154 acaatggttaacgatatcggtttcactcggtttccaccagttcaacaatatcgctaaactg 1213
Db 1151 ACAATGGTAACGATATCGGTTTCATCGGTTTCCACCAAGTCAACAATATCGCTAAACTGG 1210

Oy 1214 ttgcttccaactggtacaaatcgacgatcgaaacttccctctcgacactctgggttgcctt 1273
Db 1211 TTGCTTCCAACCTGGTACAATCGTCAGATCGAACCTTCTCTCGCACTCTGGGTGCTCTT 1270

Oy 1274 gggagttcatcccggttgatgacggttggggtgaacgtccgctgtaa 1320
Db 1271 GGGAGTTCA'PCCCGGTTGATGACGGTGGGGTGAACGTCCGCTGTAA 1317

THIS PAGE BLANK (USPTO)

RESULT 4
 XXU22962
 LOCUS
 DEFINITION Synthetic botulinum neurotoxin sero A Hc fragment (botA) gene.
 ACCESSION U22962
 VERSION U22962.1
 KEYWORDS GI:733428
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 REFERENCE 1 (bases 1 to 1338)
 AUTHORS Clayton, M.A., Clayton, J.M., Brown, D.R. and Middlebrook, J.L.
 TITLE Protective vaccination with a recombinant fragment of Clostridium botulinum neurotoxin serotype A expressed from a synthetic gene in Escherichia coli
 JOURNAL Infect. Immun. 63 (7), 2738-2742 (1995)
 MEDLINE 95310035
 REFERENCE 2 (bases 1 to 1338)
 AUTHORS Brown, D.R.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAR-1995) Douglas R. Brown, Toxinology, U. S. Army Medical Research Institute of Infectious Diseases, Bldg. 1425, Fort Detrick, Frederick, MD 21702-5011, USA
 FEATURES
 source
 1. 1338
 /organism="synthetic construct"
 /db_xref="taxon:32630"

SEQID1

1995

Query Match 98.4%; Score 1311; DB 12; Length 1338;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 16 cgtctgctgtctaccttcaactgaatacatcaagaacatcatcaatctccatcctgaac 75
 Db 15 CGTCTGCTGTCTACCTTCACCTGAATACATCAAGAATCATCAATACCTCCATCCTGAAC 74
 Oy 76 ctgcgctacgaatccaatcactcgtatcgacctgctcgtcgaagcttccaaatcaaacatc 135
 Db 75 CTGCGCTACGAATCCAATCACCTGATCGACCTGTCTCGCTACGCTTCCAAATCAACATC 134
 Oy 136 ggttctaaagttaacttcgacccgacgacagaagaatcagatccagctgttcaatctggaa 195
 Db 135 GGTTCCTAAAGTTAACTTCGATCCGATCGACAAGAATCAGATCCAGCTGTCAATCTGGAA 194
 Oy 196 tcttccaaatcgaagttaactcgaagaatgctatcgatatacaactctatgtacgaanaac 255
 Db 195 TCTTCCAAATCGAAGTTATCCTGAGAAATGCTATCGTATCAACTCTATGACGAAAC 254
 Oy 256 tcttccacctctctctggtatccggaataacttcaactcctctctctgaacat 315
 Db 255 TTCTCCACCTCTCTCTGGATCCGTATCCGGAATACTTCAACTCCATCTCTGGAACAT 314
 Oy 316 gaatacaccatcatcaactgcatggaaaacaattctggttggaagtatctctgaaactac 375
 Db 315 GAATACACCATCATCAACTGCATGGAAAACAATTCTGGTTGGAAAGTATCTCTGAACAT 374
 Oy 376 ggtgaaatcatctggaactctgcaggacactcaggaaatcaaacagcgtgtgtattcaaa 435
 Db 375 GGTGAAATCATCTGGACTCTGCAGGACACTCAGGAAATCAACAGCGTGTGTATTCAAA 434
 Oy 436 tactctcagatgatcaacatctctgactacatcaatcgtggaactctctgtaccatcacc 495
 Db 435 TACTCTCAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTCTGTACCATCACC 494
 Oy 496 aacaatcgtctgaataactccaaatctacatcaacggccgtctgactcagacagaacccg 555
 Db 495 AACATCGTCTGAATAACTCCAAATCTACATCAACGGCCGTCTGATCGACGAGAACCG 554
 Oy 556 atctccaaatctgggttaacatccacgctcttataaatacatgttcaaaactggaggtgt 615
 Db 555 ATCTCCAATCTGGGTAACTCCACGCTCTTAATAACATCATGTCAAACTGGACGGTGT 614
 Oy 616 cgtgacactcaccgctacatctggtatcaaaacttcaactctgttcgacaaagaactgaac 675
 Db 615 CGTGACACTCACCGCTACATCTGGATCAAACTTCAATCTGTTCCGACAAAGACTGAAC 674
 Oy 676 gaaaaagaaatcaaaagacctgtacgacaaccagctccaattctggtatcctgaaagacttc 735
 Db 675 GAAAAAGAAATCAAAGACCTGTACGACAACCACTCCAATCTGGGTATCCTGAAGACTTC 734
 Oy 736 tgggggtgactacacctgcagtcagcaaaacogtactacatgctgaatctgtaacgatccgaac 795
 Db 735 TGGGGTGACTACCTGCAGTACGACAAACCGTACTACATGCTGAATCTGTACGATCCGAAC 794
 Oy 796 aaatacgttgacgtcaacaatgttaggtatccgcggttacatgtacctgaaaggtccgcgt 855
 Db 795 AAATACGTTGACGTCAACAATGTAGGTATCCGCGTTACATGTACCTGAAAGGTCCGCGT 854
 Oy 856 ggttctgttatgactaccaacatctacctgaactcttccctgtaccgtggtaccaaaattc 915
 Db 855 GGTTCGTGTATGACTACCAACATCTACCTGAACCTCTCCCTGTACCGTGGTACCAAAATTC 914
 Oy 916 atcatcaagaatacgcgtctggttaacaaagacaataatcgttcgcaacaatgatcgtgta 975
 Db 915 ATCATCAAGAAATACGCGTCTGGTAACAAAGACAATATCGTTCGCAACAATGATCGTGTA 974
 Oy 976 tacatcaatgtttagttaaagaacaaagaataccgctcgtggtcccaatgctctcaggct 1035
 Db 975 TACATCAATGTTGTAGTTAAGAACAAAGATACCGTCTGGCTACCAATGCTTCTCAGGCT 1034
 Oy 1036 ggtgtagaaaagatcttctgctcgtctggaatcccggaacttggtaactctgtctcaggta 1095
 Db 1035 GGTGTAGAAAAGATCTTGTCTGCTCGAAATCCCGGACGTTGGTAATCTGTCTCAGGTA 1094
 Oy 1096 gttgtaataaatacaagaacgaccagggtatcactaacaataatgcaaaatgaatctgcag 1155
 Db 1095 GTTGTAAATGAAATCCAAAGACGACGAGGTATCACTAACAAATGCAAAATGAATCTGCAG 1154
 Oy 1156 gacaacaatggttaacgatatacgttttcaatcgggtttccaccagttcaacaatatcgctaaa 1215
 Db 1155 GACAACAATGGTAACGATATCGGTTTCATCGGTTTCCACCACTTCAACAATATCGCTAAA 1214
 Oy 1216 ctggttgccttccaactggtacaactcgtcagatcgaacgttctctcgcactctgggttgc 1275
 Db 1215

Clayton et al

THIS PAGE BLANK (USPTO)

RESULT 4
 XXU22962
 LOCUS XY 62 1338 bp DNA line: SYN 08-NOV-1995
 DEFINITION Synthetic botulinum neurotoxin serotype A Hc fragment (botA) gene, complete cds.
 ACCESSION U22962
 VERSION U22962.1 GI:733428
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM synthetic construct
 artificial sequence.
 REFERENCE 1 (bases 1 to 1338)
 AUTHORS Clayton, M.A., Clayton, J.M., Brown, D.R. and Middlebrook, J.L.
 TITLE Protective vaccination with a recombinant fragment of Clostridium botulinum neurotoxin serotype A expressed from a synthetic gene in Escherichia coli
 JOURNAL Infect. Immun. 63 (7), 2738-2742 (1995)
 MEDLINE 95310035
 REFERENCE 2 (bases 1 to 1338)
 AUTHORS Brown, D.R.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAR-1995) Douglas R. Brown, Toxinology, U. S. Army Medical Research Institute of Infectious Diseases, Bldg. 1425, Fort Detrick, Frederick, MD 21702-5011, USA
 FEATURES
 source
 1. 1338
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /gene="botA"
 /codon_start=1
 /transl_table=11
 /product="botulinum neurotoxin serotype A Hc fragment"
 /protein_id="AAA80610.1"
 /db_xref="GI:733429"
 /translation="MARLLSTFTEYIKNIINTSILNLRYESNHLIDLSRYASKINIGS
 KVNFDPIDKNQIQFLNLESSKIEVILKNAIVYNSMYENFSTFSWIRIPKYFNSISLNN
 EYTIINCMEENSGWKVSLNYGEIITLQDTQEIQRVVFQYSQMINISDYINRWIFVT
 ITNRLNNSKIYINGRLIDQKPIISNLGNIHASNNIMFKLDGCRDTHRYIWKYFNLFDM
 KELNEKEIKDLYDNQNSGILKDFWGDYLYQDKPYMLNLYDPNKYVDVNNVVGIRGYM
 YLKGPRGSMVTNNIYLNSSLYRGTKFIKKYASGNKDNVRRNDRVYINVVKNKEYR
 LATNASQAGVEKILSALEIPDVGNLSQVVVMKSKNDQGITNCKMNLNDNGNDIGFI
 GFHQFNIAKLVASNWNQRQIERSRSLTLCGSWEFIPVDDGWGERPL"
 BASE COUNT 401 a 342 c 249 g 346 t
 ORIGIN

Query Match 98.2%; Score 1302.2; DB 12; Length 1338;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 14 tggcctctaccttcactgaatacatcaagaacatcatcaatacctccatcctgaacctgc 73
 || |||||
 Db 19 TGCTGTCTACCTTCACCTGAATACATCAAGAACATCATCAATACCTCCATCCTGAACCTGC 78
 Qy 74 gctacgaatccaatcacctgatcgacctgtctcgctacgcttccaaaatcaaatcgggt 133
 |||||
 Db 79 GCTACGAATCCAATCACCTGATCGACCTGTCTCGCTACGCTTCCAAAATCAACATCGGTT 138
 Qy 134 ctaaagttaacttcgatccgatcgacaagaatcagatccagctgttcaatctggaatctt 193
 |||||
 Db 139 CTAAAGTTAACTTCGATCCGATCGACAAGAATCAGATCCAGCTGTTCAATCTGGAATCTT 198
 Qy 194 ccaaaatcgaagtattcctgaagaatgctatcgatatacaactctatgtacgaaaacttct 253
 |||||
 Db 199 CCAAAATCGAAGTTATCCTGAAGAATGCTATCGTATACAACCTCTATGTACGAAAACCTTCT 258
 Qy 254 ccacctccttctggatccgtatcccgaaatacttcaactccatctctctgaacaatgaat 313
 |||||
 Db 259 CCACCTCCTTCTGGATCCGTATCCCGAAATACTTCAACTCCATCTCTCTGAACAATGAAT 318
 Qy 314 acaccatcatcaactgcgatggaacaaattctggttggaagtattctctgaactacgggtg 373
 |||||
 Db 319 ACACCATCATCAACTGCATGGAAAACAATTCTGGTTGGAAGTATCTCTGAACCTACGGTG 378
 Qy 374 aaatcatctggactctgcaggacactcaggaaatcaaacagcgtgtgtattcaaatatct 433
 |||||
 Db 379 AAATCATCTGGACTCTGCAGGACACTCAGGAAATCAAACAGCGTGTGTATTCAAATACT 438
 Qy 434 ctcatgatcatcaatctctgactacatcaatcgctggatcttctgttaccatcaccaaca 493
 |||||
 Db 439 CTCAGATGATCAACATCTCTGACTACATCAATCGCTGGATCTTCGTTACCATCACCAACA 498
 Qy 494 atcgtctgaataactccaaaatctacatcaacggccgctctgatcgaccagaaacggatct 553
 |||||
 Db 499 ATCGTCTGAATAACTCCAAAATCTACATCAACGGCCGCTGTGATCGACCAGAAACCGATCT 558
 Qy 554 ccaatctgggttaacatccacgcttcttaataacatcatgttcaaaactggacggtgtcggtg 613
 |||||
 Db 559 CCAATCTGGGTAAACATCCACGCTTCTAATAACATCATGTTCAAACTGGACGGTTGTCTGTG 618
 Qy 614 aactcaccgctacatctggatcaaatcttcaatctgttcgacaagaactgaacgaaa 673
 |||||
 Db 619 AACTCACCCTACATCTGGATCAAACTTCAATCTGTTCGACAAGAAGTGAACGAAA 678
 Qy 674 aagaaatcaagacctgtacgacaaccagtcgaattctggtatcctgaaagacttctggg 733

Clayton et al
 1995

SEQIDS

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 17:10:08 : Search time 7254.67 Seconds
(without alignments)
3842.235 Million cell updates/sec

Title: US-09-611-419A-1

Perfect score: 1332

Sequence: 1 gaatcgaacagctgcgtc.....aacgtccgcgtgaagaattc 1332

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_hlg:*
- 3: gb_in:*
- 4: gb_ov:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sy:*
- 12: gb_un:*
- 13: gb_un:*
- 14: gb_vl:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_or:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_pl:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vl:*
- 30: em_hlg_hum:*
- 31: em_hlg_inv:*
- 32: em_hlg_other:*
- 33: em_hlg_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No. Score Match Length DB ID Description

1	1311	98.4	1330	6	AR000029	AR000029 Sequence
2	1311	98.4	1330	6	AR169140	AR169140 Sequence
3	1311	98.4	1330	6	AX036243	AX036243 Sequence
4	1311	98.4	1338	12	XX022862	U22962 Synthetic b
5	1311	98.4	1402	6	AR000030	AR000030 Sequence
6	1311	98.4	1402	6	AR169141	AR169141 Sequence
7	1311	98.4	1402	6	AX036246	AX036246 Sequence
8	727.6	54.6	4292	6	CBROTAG	X52066 Clostridium
9	727.6	54.6	4835	1	CLONEUR	M30196 C.botulinum
10	726.4	54.5	3891	6	AR000031	AR000031 Sequence
11	726.4	54.5	3891	6	AR169142	AR169142 Sequence
12	726.4	54.5	3891	6	AX036248	AX036248 Sequence
13	662	49.7	4067	1	CBNTOX	X73423 C.botulinum
14	423.2	31.7	1313	6	AF51281	AF51281 Synthetic
15	421.6	31.7	1313	6	AF58946	A58946 Sequence 6
16	227.8	17.1	4073	1	CBRONTF	X68262 C.barrati ge
17	227.2	17.1	1293	6	A58945	A58945 Sequence 5
18	227.2	17.1	4199	1	CLOBONT	L35496 Clostridium
19	227.2	17.1	4209	1	CBRONTFG	X81714 C.botulinum
20	195	14.6	3835	1	AB037704	AB037704 Clostridi
21	195	14.6	3835	1	AB037705	AB037705 Clostridi
22	195	14.6	3835	1	AB037706	AB037706 Clostridi
23	195	14.6	3835	1	AB037707	AB037707 Clostridi
24	195	14.6	3835	1	AB037708	AB037708 Clostridi
25	195	14.6	3835	1	AB037709	AB037709 Clostridi
26	195	14.6	3835	1	AB037710	AB037710 Clostridi
27	195	14.6	3835	1	AB037711	AB037711 Clostridi
28	195	14.6	3835	1	AB037712	AB037712 Clostridi
29	195	14.6	3835	1	AB037713	AB037713 Clostridi
30	195	14.6	3835	1	AB037714	AB037714 Clostridi
31	195	14.6	4030	1	CBNTTE	X62683 C.botulinum
32	193.4	14.5	4017	6	AX088262	AX088262 Sequence
33	193.4	14.5	4017	6	CBNTTOXE	AB039264 Clostridi
34	190.2	14.3	3838	1	AB039264	X62088 C.botulinum
35	190.2	14.3	3838	1	CBNTTE	M92906 Clostridium
36	180.8	13.6	9234	1	CLONEUROXF	Y13631 Clostridium
37	180.8	13.6	9234	1	CBY13631	I28431 Sequence 3
38	166.4	12.5	1359	6	I28431	A49987 Sequence 4
39	166.4	12.5	3712	6	A49987	A37074 Sequence 17
40	166.4	12.5	3754	6	A37074	A42478 Sequence 6
41	166.4	12.5	3754	6	A42478	A37075 Sequence 18
42	166.4	12.5	3769	6	A37075	A49988 Sequence 5
43	166.4	12.5	4366	6	A42484	A42484 Sequence 12
44	166.4	12.5	4366	6	A42484	A42481 Sequence 9
45	166.4	12.5	4378	6	A42481	

ALIGNMENTS

RESULT 1

LOCUS AR000029 1330 bp DNA

DEFINITION Sequence 22 from patent US 5736139.

ACCESSION AR000029

VERSION AR000029.1 GI:3962560

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1. (bases 1 to 1330)

AUTHORS Kink,J.A., Thalley,B.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.

TITLE Treatment of Clostridium difficile induced disease

JOURNAL Patent: US 5736139-A 22 07-APR-1996;

FEATURES

source 1.1330

BASE COUNT 400 a 339 c 246 g 345 t

ORIGIN

Query Match 98.4%; Score 1311; DB 6; Length 1330;

Best local Similarity 100.0%; Pred. No. 0;

[illegible]

QY	1096	ggtgtaagaaatccaagagacagcagggtatcacttaaaaaatgaaaaatgatctgcag	1155
QY	1087	gtttgttaataaataccaagaacacacagggatctacttaacaaaatgaaatgatctgcag	1146
QY	1156	gacacaaatggtaaagatalcgcgttttcaatcgcgttttccaccagtttcaacatalcgcataa	1215
Db	1147	gacacaaatggtaaagatalcgcgttttcaatcgcgttttccaccagtttcaacatalcgcataa	1206
QY	1216	ctgtgttcttccaaactgylacaaatcgtaagatcgaaacgltctcttcgaacttgggttgc	1275
Db	1207	ctgtgttcttccaaactgylacaaatcgtaagatcgaaacgltctcttcgaacttgggttgc	1266
QY	1276	tcttggaattatcccggtttgatagcagtttggttggaacgttcgcgtgtaa	1326
Db	1267	tcttggaattatcccggtttgatagcagtttggttggaacgttcgcgtgtaa	1317
RESULT	2		
LOCUS	AR169140	1330 bp	DNA
DEFINITION	Sequence 22 from patent US 6290960.		Linear
ACCESSION	AR169140		
VERSION	AR169140.1	GI:17906909	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	1 (bases 1 to 1330)		
AUTHORS	Kink, J.A., Thalley, B.S. and Stafford, D.C.		
TITLE	Vaccine and antitoxin for the treatment of C. difficile disease		
JOURNAL	Patent: US 6290960-A 22 18-SEP-2001;		
FEATURES	Location/Qualifiers		
SOURCE	1..1330		
BASE COUNT	400 a 339 c 246 g 345 t		
ORIGIN			
Query Match	98.4%: Score 1311: DB 6: Length 1330:		
Best Local Similarity	100.0%: Pred. NO. 0: Mismatches 0: Indels 0: Gaps 0:		
Matches 1311: Conservative	0: Mismatches 0: Indels 0: Gaps 0:		
QY	16	ctgtcgtcttccacttccatgtaataacaaacatcaatcaatcaatccatccatccgaac	75
Db	7	cgctgctgcttacccttccatgtaataacaaacatcaatcaatcaatccatccatccgaac	66
QY	76	ctgtgctacgaatccaatcaatcaatctgtaacgtctgtctgtaacgttccaaatcaatc	135
Db	67	ctgtgctacgaatccaatcaatcaatctgtaacgtctgtctgtaacgttccaaatcaatc	126
QY	136	ggtttcaagtttaacttgcgaatcgcgaatcgaaagaatcaatcaatcagctgttcaatcgaa	195
Db	127	ggtttcaagtttaacttgcgaatcgcgaatcgaaagaatcaatcaatcagctgttcaatcgaa	186
QY	196	tcttccaaatcgaaatltacttgaagaatgtctatcgltatacaactctatgtgaagaac	255
Db	187	tcttccaaatcgaaatltacttgaagaatgtctatcgltatacaactctatgtgaagaac	246
QY	256	tcttccaaatcgaaatltacttgaagaatgtctatcgltatacaactctatgtgaagaac	315
Db	247	tcttccaaatcgaaatltacttgaagaatgtctatcgltatacaactctatgtgaagaac	306
QY	316	gaatcaacatcaatcaatcgaaatgtgaagaatgtctatcgltatacaactctatgtgaagaac	375
Db	307	gaatcaacatcaatcaatcgaaatgtgaagaatgtctatcgltatacaactctatgtgaagaac	366
QY	376	ggtgaatcatctggaatctggaatgtgaagaatgtgaagaatgtgaagaatgtgaagaat	435
Db	367	ggtgaatcatctggaatctggaatgtgaagaatgtgaagaatgtgaagaatgtgaagaat	426
QY	436	tactctcagatgatacaaatcttgaactatacaatcgctgtgatacttgcgttcaacacc	495

Qy	676	gaaaaagaatcaaaagccgtgtacgaacaacacgttccatcttgtagtcccgaaagacttc	735
Db	667	gaaaaagaatcaaaagccgtgtacgaacaacacgttccatcttgtagtcccgaaagacttc	726
Qy	736	ctgggtgtactacgtcgtacgaacaacccgtaactatgctgtaactgtgtaacatccgaac	795
Db	727	tgggggtactacgtcgtacgaacaacccgtaactatgctgtaactgtgtaacatccgaac	786
Qy	796	aaatacgttgcgtcgaacaatgtatgattccgggttatactgtacactgaaagctccggt	855
Db	787	aaatacgttgcgtcgaacaatgtatgattccgggttatactgtacactgaaagctccggt	846
Qy	856	ggttcgtgtatgacttacaacatctacttgaactcttcctcttccatccgtgttacaatttc	915
Db	847	ggttcgtgtatgacttacaacatctacttgaactcttcctcttccatccgtgttacaatttc	906
Qy	916	atcatalcaagaataacgcgtctgtgtacaaagaacataatcgtctgcaacaatagtatcgtga	975
Db	907	atcatalcaagaataacgcgtctgtgtacaaagaacataatcgtctgcaacaatagtatcgtga	966
Qy	976	tacatcaatgtgtgtatgtatgaagaacaataaccgctctgtgctacccaatgtcttcagact	1035
Db	967	tacatcaatgtgtgtatgtatgaagaacaataaccgctctgtgctacccaatgtcttcagact	1026
Qy	1036	gggttgaagaagaatcgtgtctgtctgtcgtgaatcccgagcgtgtgtatcgtgtcgaagta	1095
Db	1027	gggttgaagaagaatcgtgtctgtctgtcgtgaatcccgagcgtgtgtatcgtgtcgaagta	1086
Qy	1096	gttgttaatgaatacccaagaacacgaaggtatcactatacaaatgcaaaatgtaactcgcag	1155
Db	1087	gttgttaatgaatacccaagaacacgaaggtatcactatacaaatgcaaaatgtaactcgcag	1146
Qy	1156	gacaacaatggttaacgatatcgtgttcatcgtgttccacagttcaacaatatcgtctaa	1215
Db	1147	gacaacaatggttaacgatatcgtgttcatcgtgttccacagttcaacaatatcgtctaa	1206
Qy	1216	ctggttgccttccacactggttacaatcgtcagatcgaacgtcttcctctcgaactgggttc	1275
Db	1207	ctggttgccttccacactggttacaatcgtcagatcgaacgtcttcctctcgaactgggttc	1266
Qy	1276	tcttggaggttcatcccggttgaatgaacggttgggtgaacgtcgcgtgttaa	1326
Db	1267	tcttggaggttcatcccggttgaatgaacggttgggtgaacgtcgcgtgttaa	1317
RESULT 4			
LOCUS	XU22962	1338 bp	DNA linear SYN 08-NOV-1995
DEFINITION	Synthetic botulinum neurotoxin serotype A Hc fragment (botA) gene, complete cds.		
ACCESSION	U22962.1 GI:733428		
VERSION	U22962		
KEYWORDS	synthetic construct, artificial sequence.		
SOURCE	1 (bases 1 to 1338)		
ORGANISM	Clayton, M.A., Clayton, J.M., Brown, D.R. and Middlebrook, J.L.		
REFERENCE	Protective vaccination with a recombinant fragment of Clostridium botulinum neurotoxin serotype A expressed from a synthetic gene in Escherichia coli		
AUTHORS	Infect. Immun. 63 (7), 2738-2742 (1995)		
JOURNAL	J95310035		
MEDLINE	2 (bases 1 to 1338)		
REFERENCE	Brown, D.R.		
AUTHORS	Direct Submission		
TITLE	Submitted (17-MAR-1995) Douglas R. Brown, Toxicology, U. S. Army Medical Research Institute of Infectious Diseases, Bldg. 1425, Fort Detrick, Frederick, MD 21702-5011, USA		
JOURNAL	Location/Qualifiers		
FEATURES	1. 1338		
SOURCE	/organism="synthetic construct"		
	/db_xref="taxon:32630"		

Db	675	GAAAAAGAAATCAAGACCTGTACGACAAACCAAGTCCATTCTCGTATCTCTGAAACATCTTC	734
OY	736	tggggtgacctaccgcycaagtaacgaacaaccgctaactacatgcttgaatctgtaacgtccgaac	795
Db	735	TGGGGTGCATCTACTCTCAGTACGACAAACCGTATCTCAATGTGAAATCTGTACATCTCGAAGC	794
OY	736	aaatcgtttgacgttaacaaatgtagtatacccggtttacatgtaactgtctgaaggttcgcgt	855
Db	735	AAATACGTTTGACGTCAACAATGTAGGTATCCGGGTTTACATGTACCTGAAGAGTCCGCGT	854
OY	856	ggttcgttatactaccacaacatctaccctgaacctctccctgtaccgltgtaaccaatlc	915
Db	855	GGTTCTGTATTAGACATACCACAATCTACCTTGAACTTCCCTGTACCGTGGTACCAAAATTC	914
OY	916	atctatcaagaatacgcgtctgtgttaacaagacaatatcgtttcgcaaaatgatacgttga	975
Db	915	ATCATCAAGAAATATGCGCGTCTGGTAAACAAGGCAATATCTGTCGCAACATATATGTGTGA	974
OY	976	tacatcaatctgttgtaagtaagaacaagaataacgcgttcggctaccatgctctctcaagct	1035
Db	975	TACATCATATGTTGTAGTTAAAGAACAAAGATATCCGTTGCGTACCAATGCTTCCAGGCT	1034
OY	1036	ggttgaagaagaatcttgtctgtcctctgtgaatacccgagcgttggtaactgtgtctcaagta	1095
Db	1035	GGTGTAGAAAGATCTTGTCTGTCTCTGTGAATAACCGGAGCGTGTGTAAATCTGTCTAGGTA	1094
OY	1096	gtctgaatgtaaatccaaagaacgacccaaggtatcacatacaataatgtcaaaatgtaatctgcag	1155
Db	1095	GTTTGAATGAATCCAAAGACGACCAAGGATATCATTAACAATGTGCAAAATGAAATGTACGAG	1154
OY	1156	gacaaacaatgtgaacgatatatcggtttcaatcggttccaccacagtttaacaataatctgctaa	1215
Db	1155	GACAAACAATGGTAACGATATCGGTTTCAATCGGTTTCCACCAAGTTCAACAATATTCGTATAA	1214
OY	1216	ctggttgcttccaactggtlacaatcgtlcaagatcgaaagcttccctctgcacactctgggttgc	1275
Db	1215	CTGGTGTGCTTCCAACTGTAATGTGTAATGTGAGATCGAAGACGTTCTCTCGCACTCTGGGGTTC	1274
OY	1276	tcttggaggtatcatccggttgatataagttggggttgaaagcgcgctgtaa	1326
Db	1275	TCTTGGAGATTCATCCCGGTTCATACCGTTTGGGTGAACGTCGCCGTGTAA	1325

|||||
Db 1219 GACAAACATGGTAACGATATCGGTTTCATCGGTTCCACAGTTCAACATATCGCTAA 1278
Qy 1216 ctgtgtcttcacacatgtagcaatcgtccagatcgaaacgtcttcctcgcactctggtgtgc 1275
Db 1279 CTGGTTGCTTCCAAAGTGATCAATCGTCAGATGGAACGTTCTCTCGCACTCGGGTTGC 1338
Qy 1276 tcttggagatcattcccggttgtagacggttgggtgaaacgtccgtctgtaa 1326
Db 1339 TCTTGGAGATTCAATCCCGGTTGATGACGGTTGGGGTGCAACGTCGCTGTAA 1389
RESULT 6
AR169141 1402 bp DNA linear PAT 17-DEC-2001
LOCUS AR169141
DEFINITION Sequence 25 from patent US 6290960.
ACCESSION AR169141
VERSION AR169141.1 GI:11906911
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1402)
AUTHORS Kink,J.A., Thalley,B.S. and Stafford,D.C.
TITLE Vaccine and antitoxin for the treatment of C. difficile disease
JOURNAL Patent: US 6290960-A 25 18-SEP-2001;
FEATURES
source 1..1402
Location/Qualifiers
BASE COUNT 420 a 360 c 260 g 362 t
ORIGIN
Query Match 98.4%; Score 1311; DB 6; Length 1402;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 556 atctcaactctgggttaacatccacgcctctcaataacatcgttcaacctggacggttgc 615
Db 619 ATCTCAATCTGGGTAAACATCCACGACTTCTTAATTAACATGTTCAAACTGGAGCGTTGT 678
Qy 616 cgtgacacacgcctacatcctgatatcaataactacatcgttctcgacaagaactgaac 675
Db 679 CGTGACACTACCCGCTTACATCTGGATCAAAATACTTCAATCTGTTCCGCAAAAGAACTGAAC 738
Qy 676 gaaagaagaatcacaagaccctgtacgacaacacgltcccaatctgtgltccctgaaagacttc 735
Db 739 GAAAAAGAAATCAAAAGACCTGTACGACACACGATCCAAATCTGGTATCTCGAAAGACTTC 798
Qy 736 tgggtgtactactcctgcagatcagaacacgctactacatctgtgaatcgttcagatccgaac 795
Db 799 TGGGGTACATCCCTCGATCGACAAACCGGTACTACATGTTGATGATCGATCCGAAC 858
Qy 796 aaatcagctgacgtaacaatgtagtaccgcggtgtacatgtaacgtaagagtcgcgt 855
Db 859 AAATACGTTGACGTCACAAATGTAAGTATCCGGGTTACATGTAACCTGAAGGTCGCCGT 918
Qy 856 ggtctgtttagtaccacacatcctacatccttccctgtaacggttgtaaccaatlc 915
Db 919 GGTCTGTATAGACTACCAACATCTACCTGAACTCTCCCTGATCCGTTGATACCAATTC 978
Qy 916 atcaatcaagaataacgctgtgtgtaacaaggaacaatctgttcgacaacatgacgtgta 975
Db 979 ATCATCAAGAAATACCGGCTGTGTACCAAGACAAATATCTGCGCAACATGATCGTGA 1038
Qy 976 taactcaatgttgtagttagtaagaagaataacacgtctggtctaaccaatctgctcagac 1035
Db 1039 TACATCAATGTTGTAGTTAAAGAACAAAGAAATACCGTGTGGCTTACCAATGCTTCTCAGGCT 1098
Qy 1036 ggtgtagaagaatctgtctgtcctgtgaaatcccggaacgttggtaactctgtctcagta 1095
Db 1099 GGCTGTAAAAAGATCTGTCTGCTGTGGAATCCCGGACGTTGGTATCTGTCTCAGGTA 1158
Qy 1096 gttgtatgaataatccaaagaacgacaggtgataactaacaagaatgaatcgtcag 1155
Db 1159 GTTGTATGAAATCCAAAGAACGACCAAGGATATCTTAACAAATGCAAAATGCAATCTCAG 1218
Qy 1156 gacacaatgttgaacgatacgtgttaccaggttccaccaggttcaacaatatcgctaa 1215
Db 1219 GACAAATATGTAACGATATCGGTTTCAATCGGTTTCCACAGTTCAACAAATATCGCTAA 1278
Qy 1216 ctgtgtcttcacacatgtagcaatcgtccagatcgaaacgttcttcctcgcactctggtgtgc 1275
Db 1279 CTGGTTGCTTCCAAAGTGATCAATCGTCAGATGGAACGTTCTCTCGCACTCGGGTTGC 1338
Qy 1276 tcttggagatcattcccggttgtagacggttgggtgaaacgtccgtctgtaa 1326
Db 1339 TCTTGGAGATTCAATCCCGGTTGATGACGGTTGGGGTGCAACGTCGCTGTAA 1389
RESULT 7
AX036246 1402 bp DNA linear PAT 16-NOV-2000
LOCUS AX036246
DEFINITION Sequence 25 from Patent EP1041149.
ACCESSION AX036246
VERSION AX036246.1 GI:11225864
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 1402)
AUTHORS Kink,J.A., Firca,J.R., Padhye,N.V., Thalley,B.S., Stafford,D.C. and Williams,J.A.
TITLE Vaccine and antitoxin for treatment and prevention of C. Difficile disease
JOURNAL Patent: EP 1041149-A 25 04-OCT-2000;
FEATURES
source 1..1402
Location/Qualifiers
/organism="synthetic construct"

CDS

/db_xref="taxon:32630"
1..1389
/note="Synthetic"
/codon_start=1
/transl_table=1
/protein_id="CAC16479.1"
/db_xref="GI:11225865"
/translation="MGHHHHHHHHSSGHIEGRHMAARLSTFEYIKNIITSI
LNRYSNHLIDLSRYASKINIGSKVNDPIIDKNOIOLEFNESKIEYLKNAIYNS
MYENFSFMRIRIKYFNSISLNREYITINENNSGKYLNGEITWTLODVEIK
ORVFKTSOMINISDYINRMLFVITITNRLNNSKIYINGRLIDQFISLNINHASNN
IMFKLDGCRDHRH IWLKYFMLEFKLEKEIKDLIDQNSGILKIDGDIYLODKP
YMLNLYDPKRYVDVNNVNGIRGYMLKGRGVSMTNTNLYLSNLSYRGFKFIKKAAG
NKDNIVRNDRVYINVVYKNEKRYLATNASOAGVEKILSALEIPDVGMLSOVVMKSK
NDGITNKCKMNLQDNNNGNDIGFIFGHQFNFIATLVANMWNRQERSRLLGCSMEF
IPVDGMCERL"

ASE COUNT 420 a 360 c 260 g 362 t
RIGIN

Query Match 98.4%; Score 1311; DB 6; Length 1402;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

16 cgtctcgtctctacacctctactgaatacacaagaacatcatcaactcctcgaac 75
|||||
79 CGTGTGCTGTCTACCTTCAGTAATACATCAAGAACATCATCAATCCATCCGTGAAC 138
|||||
76 ctgagctacgaatccaatcaactctgaactgtctcgtctgaactcctcaaatc 135
|||||
139 CTGCGCTAGGAATCCAAATCACCCTGATCGACCTGTCTCGTACGGCTCCAAATCAACATC 198
|||||
136 ggtctaaagttaactctcgatccgacatcgagaacatcagatccagctgttcaactcgtga 195
|||||
199 GGTCTTAAGTTAACTTCGATCCGATCCGACAGAAATCAGATCCAGCTGTTCATCTCGAA 258
|||||
196 tcttccaaatcgaagtatccttgaagaatgtctatcgtatatacaactctatgtcgaagaac 255
|||||
259 TCTTCCAAATCGAAGTATTCTCGAAGAAATGCTATCGTATACACTCTATGTAACGAAGAAC 318
|||||
256 tttctcaactccttctgatacctgtatcccgaaatcttaactcctcctctgaagaac 315
|||||
319 TTCTCCACCTCTCTTGTGATCGTATCCGAAATACTTCATCTCATCTCTGTGAACAAT 378
|||||
316 gaatacacatcatcaactctgaagaagaacatctcgtgtgaagaatctctgaactac 375
|||||
379 GAATACACCATCATCACTCGATGGAAGAAACATTTCTGTGGAAGTATCTGTGAATGAC 438
|||||
376 ggtgaatcatctggaactctcgaagacactcaggaatcaaacagcgtgtgtatlatcaaa 435
|||||
439 GGTGAATCATCTCGACCTCTGACAGACACTCAGAAATCAACACGCTGTGTATTCAAA 498
|||||
436 tactctgagatgataaactctctgaactatacatcactcgtgtatcttctgttaccatcac 495
|||||
499 TACTCTCAGATGATCAACAATCTCTGATPACATCAATCGCTGATCTTCTGTACCATCACCC 558
|||||
496 aacaaatcgtctgaactcaactcaaaatctatacatcaacgagcgtctgatcagagaagaacg 555
|||||
559 AACATCGTCTGAATTAACGCCAAATCTATCATCAACGCGCTGATGACCAAGAAACGG 618
|||||
556 atctccaactctgggtgaatcaatcagcctcttaataacatcatgttcaaacctgaagctgt 615
|||||
619 ATCTCCATCTGGGTGAATCATCCAGCTTCTATTAACATCATGTTCAAACTGGAAGGTTGT 678
|||||
616 cgtgacatcaaccgtctatcatctgtgatcaaatctcaatctgtctgaagaagaactgaaac 675
|||||
679 CGTGACATCAACCGCTATCATCTGATCAAAATCTTCATCTTGTGACAAACACTGGAAC 738
|||||
676 gaaaagaatcaaaagactctgaagaacaacagctcaactctcgtctatctgaagaacttc 735
|||||
739 GAAAAAGAAATCAAAAGACTCTGTACACAAACGCTCAATTTCTGGTATCTTGAAAGACTTC 798
|||||
736 tgggggtgactactcgtcagtaacagaacacgtactatcatgtcgtatctgtacgataccgaac 795

|||||
Db 799 TGGGGTACTACTGACGTACGACAAACCGTACTACATCTGTAATCTGTAACGTCGAAC 858
|||||
796 aaatacgttgaagctcaacaatgtagttatcccggtgttaactgtactgaagaagctcgt 855
|||||
859 AATATAGCTTGAGCTCAACATGATGATGATCCGGGTACATGATCACTGTAAGGTCCGCT 918
|||||
856 ggttcgttatctactcaacaactctacactcgaactctccctcttacggtgtgtacaaatc 915
|||||
919 GGTTCGTATGATGATCAACATCTACTGAACTCTTCCCTGTACCGTGTGTAACCAATTC 978
|||||
916 atcatcaagaanaatcagctcgtctgttaacaagaacatcgttctgcgaacaatgatctgtga 975
|||||
979 ATCATCAAGAAATATACCGCTGCTGTATCAAGCAATATGCTTCGCAACATGATCGTGA 1038
|||||
976 tacatcaatgttgtgttgaagaacaagaatcagctcgtctgtcctacatgtcttcagct 1035
|||||
1039 TACATCAATGTTGTGATTAAGAAACAAAGATACCGTCTGCTACCAATCTTCTCAGGCT 1098
|||||
1036 ggtgtagaagaatcctgtctgtctgtgaatacccgagcgttgttaactgtctcagcta 1095
|||||
1099 GGTGTGAAAGATCTTGTCTGCTGTGAAATCCCGGACGTTGTATCTGTCTCAGTA 1158
|||||
1096 gttgttaatgaatccaagaacgacagggatcactaacaatgtcaaatgaatctgcag 1155
|||||
1159 GTTGTATGAATCCAAAGACGACGAGGTATCACTAACAAATGAAATGATCTGACG 1218
|||||
1156 gacaacaatgtgaagatctgtgttctcagctgttccacagcttcaacaatctcgttaa 1215
|||||
1219 GACAAACAAATGTAACGATTCGGTTTCATCGGTTTCCACCACTTCAACAAATTCGTTAA 1278
|||||
1216 cgtgtctctcgaactcgtgtacaaatcgcagatcgaacgttctcctcgcactcgtgtgc 1275
|||||
1279 CTGTTGCTTCCACATCGTATCAATCTGATGATGAAAGTTCTCTGCACTCTGGGTTGC 1338
|||||
1276 tcttggagatcaatcccggtgtatgaagcttggggtgaacgtctcgtgtaa 1326
|||||
1339 TCTTGGAGCTTCAATCCGGGTGATGACGGGTGGGGAACGTCCTCGTTAA 1389
|||||

RESULT 8
CBOTAG
LOCUS CBOTAG 4292 bp DNA linear BCT 12-SEP-1993
DEFINITION Clostridium botulinum botA gene for type A neurotoxin.
ACCESSION X52066.X52088
VERSION X52066.1 GI:40381
KEYWORDS botA gene; neurotoxin; secreted protein.
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 4292)
AUTHORS Minton,N.P.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-1990) Minton N.P., PHUS Centre for Applied
Microbiology & Research, Molecular Genetics Group, Division of
Biotechnology, Porton Down, Salisbury SP4 0UG Wiltshire, U K
2 (bases 1 to 4292)
REFERENCE Thompson,D.E., Brehm,J.K., Oultrem,J.D., Swinfield,T.J.,
Shone,C.C., Atkinson,T., Welling,J. and Minton,N.P.
TITLE The complete amino acid sequence of the Clostridium botulinum type
A neurotoxin, deduced by nucleotide sequence analysis of the
encoding gene
JOURNAL Eur. J. Biochem. 189 (1), 73-81 (1990)
MEDLINE 90235864
FEATURES
source Location/Qualifiers
1..4292
/organism="Clostridium botulinum"
/strain="sub sp. type A, NCRC2916"
/db_xref="taxon:1491"
16..64
/note="dyad symmetry"
19..24
promoter

REFERENCE	TITLE	JOURNAL	COMMENT
1 (bases 1 to 4835)	Bacteria; Firmicutes; Bacillus/Clostridium; Clostridium.	Binz, J., Kurazono, H., Wille, M., Frevet, J., Wernars, K. and Niemann, H.	The complete sequence of the botulinum type A neurotoxin and its comparison with other Clostridial neurotoxins
90264400	J. Biol. Chem. 265, 9153-9158 (1990)		
by H.Niemann, 29-NOV-1989.			
FEATURES	Source	Location/Qualifiers	
misc_feature	1..4835	/organism="Clostridium botulinum"	
misc_feature	/db_xref="taxon:1493"		
misc_feature	240..>4835		
misc_feature	/product="neurotoxin mRNA"		
misc_feature	344..349		
misc_feature	/note="ribosome binding site"		
misc_feature	358..4248		
misc_feature	/note="neurotoxin"		
misc_feature	/codon_start=1		
misc_feature	/transl_table=1		
misc_feature	/protein_id="AA033262.1"		
misc_feature	/db_xref="GI:144865"		
misc_feature	/translation="MPEVNRQFNYKDPVNGVDIAYIKIPNAGOMQPYAKRINKHWI IPEDFTPEEGDLNPPPEAKQPVSYSDYSTLSTDNEDKNTLKGATLFEERYSTD LGRMLTISVIRGIEPMFGSGTIDELKVIDNCINIVIOPODSYREEELNVIIGSADI IOPEKSGFGEHLVNLTRNGSGTQYIRSPDPFGPEESLEVDNPLLGAKRATDPA VTLAHELHAGHGRVLYGAIANPNRERKNTAYAYEMSLVSEPELRITFGCHDAFIDS LOEHEPRLIYKTRKDIASLTLNRKASIVGTASIAQIYKNVKEKYLLESDTSGKFSVD KLFNDLKYKMLTEIYEDNFEKFEKYNRRYTLNFDCAVRIINVPVNTIYDGMAL RNNLNLKNGMONTNINNMNFTKLNFTGFEELKLVGIIITSKRSLDKGYNKL NDLICIKVNMNDLFPSPEEDNTNDLNGEETSPNTEAEENISDLIOOYITFNF DNDEPNSIENTISDITGOLEIMPINERPNRGKYEIDKTYMPEYLRABOEFHRSRI AKTNSVEALINPRVRYFPSSPYKRVKNTAEAMLGVBEOLVYPTDTSVSTT DKADITITITIPISPALNTGMILKTDVPGALIFSGAVILDEFIPELAIPELIFALV SYIANKLVVQTTIDNALSKRNEKWEYKTIYVNMIAKVQTQIDLIRKKKEALENOA EATGAIINVOYNQTEEEKNNINFINIDLSKLSINRKAMININKLOCSSYLKN SMAYGVKRLDEPDFASLDAKLKITYNRGTLIGOVNRDKDVNNTLSTDIPOLSKY VDNRLSTFEVYIKNTINRSLNLRRESNHLIDLSRYASINIGSKVNPDPDKNOI QLPRESKEIEVILKNAIYVNSMEKNSSTSPRIPIKPYFSSISLNRYTITINCENNS GMAVSLVSGEITWTLQDTEIKORVVEYQSMINIQYINRWIEVITVNNRNLNSKLY INGLIDQKRPISNLGNHASNNINMIFKLDGCRDRIYIWIYFNLDELNEKEIKDLY DNQNSNGSLKDPMDGYLDQRYPMALYDYPKVDVANNVIGRYVNLKCPGQSVMT NIYLSNLYRGYKTIKKYASGNKNDVRRNDRYINVVYKNEYRLATNASAGVKK ILSLTEIPDGNLSQVYVMSKNDQGITNCKNMLDNNNGNDIGFISGHQPNIAKLY ASNYNRQIRSSSTLDCSWEFIYVDGCGERPL"		
misc_feature	4400..4432		
misc_feature	/note="potential terminator: putative"		
misc_feature	1934 a 517 c 756 g 1628 t		
BASE COUNT	ORIGIN		
Query Match	54.6%; Score 727.6; DB 1; Length 4835;		
Best Local Similarity	72.2%; Pred. No. 6.5e-177;		
Matches 946; Conservative	0; Mismatches 364; Indels	0; Gaps	0;
23	tgctacctgaatgaataacatcaagaacatcatcaataacatcctgaactgctg	82	
2945	TATCTACATTTACTGCAATATATTAAGATATATTATTAATCACTCATATTGAATTTAAACAT	3004	
83	acgaatcaatcaatcagctgacacgctgctgcctgaactccaaataacatcggttca	142	
3005	ATGAAAGTAATCATTTATATAGACTTATATAGTATAGCTACAAAATTAATTTGGTAGTA	3064	
143	aagtaactcgatcgatcgatcgacaagaatcagatcagctgltcaatcgtgaattctcca	202	
3065	AAGTAATTTTGCATCAGATAGATAAATTCAAATTCATTAATTAAATTAGAAAGTAGTA	3124	
203	aatcgaagttatcctcgaagaatgctatcgatatacaactcatgtacgaactctcca	262	

D	3125	AAATTGAGGTAATTTTAAAAAATGCTATGTGTAATATATGTAATGTAATGTAATTTTAACTA	318
Q	263	ccctctcttgatccgcatcccgaaatctcaactccatctctctgaaacaatgaaataca	322
D	3185	CTACCTTTTGAGTAAGAAATTCCTAAGATATTTTAAACGTATTAAGTCTAAATTAATGAATATA	3244
Q	323	ccatcatcaactcgtcaygaaacaattctggttggaaagatctctgtgaactcgggtgaa	382
D	3245	CAATTAATTAATTTGATATGGAATAATTAATTCAGATGGAAGATACACTTAATTTTGCTGAAA	3304
Q	383	tcactctggaacttcgacgaacactcagaaatcaacaagcgtttgattcaaatctctc	442
D	3305	TAACTGCACTTTACAGGATACCTGAGAAATTAACAAAGACTAGTTTAAATATCACTC	3366
Q	443	agatgatcaaatctctcgactatcaatcgaatcgtcgatctctgttaccatcaccaacaatc	502
D	3365	AAATGATTAATTAATTCAGATTAATTAATTAACAGATGATTTTGTAACTATCACTAATATA	3422
Q	503	gtctgataactccaanaattcacaaagcgccgtctgtatgacacgaagaacgcatctca	562
D	3425	GATTAAATTAACCTCAAAAATTTATATAATGGAAGATTAATAGATCAAAAACCAATTTTAA	3484
Q	563	atcgggtaacaatccacgcttccataaactaatgttcaaacctggacggttctgtgca	622
D	3485	ATTTAGGTAATTAATTCACGTACGTAGTAATTAATTAATGTTTAAATAGATGTTTAAAGATA	3544
Q	623	ctcacgcctacatctgtgatccaatctcaatctgtctcgacaagaactgaaacgaaag	682
D	3545	CACATGATTAATTTTGATTAATTAATTTTAACTTTTGTATTAAGAAATTAATTAATGAAGAAG	3604
Q	683	aaatcaagaacctgtacgaacaaccggtccaattctggtatcctgaaagaactctgggtg	742
D	3605	AAATCAAGAATTTATTAATGAATTAATCAATCAAAATTCACAGTATTTTAAAGACTTTTGCGGTG	3664
Q	743	actaccgcgactgacgaacaaccgtaactaacgtgtgatctgtacgaactccgacaacaatcg	802
D	3665	ATTATTTTACAAATTAATGAATTAACCACTACTATATGTTAAATTTAATATGATCCAAATTAATATG	3722
Q	803	ttgacgtcaacaatgtagtataatccgcggttatactgtacactgtaaaaggttcgcggttctg	862
D	3725	TGCAATGAATAATTAATGTAGTAATTAAGAGTATATGTAATCTTAAAGGCGCTAGAGGTACGG	3784
Q	863	ttatgataccaacatctacactgaactcttccctgtacccgttggttaccaaattcatcatca	922
D	3785	TAAAGCTACAAACATTTATTTTAAATCAAGTTTGTAATGGGGGACAAAATTTATTAATTA	3844
Q	923	agaatacgcgtctgctgaaacaagaacataatcgttcgacaacatgatactgtatcatca	982
D	3845	AAAAATATGCTCTCGGAATAATTAAGATTAATTAATTTTGAATAATATGATCCTGTAATTAATA	3904
Q	983	atgtgtgagttaagaaacaagaataacgctctggtctacaatctcttcacggtctgtgtag	1042
D	3905	ATGTAGTAGTAAATAAATAAAGAAATTAATAGTTTACACTATATGCAACACAGCGACGCTAG	3966
Q	1043	aaaagatctgtctgctctgynaatcccggaacgttggtaactcgtctcaggtagttgtaa	1102
D	3965	AAAAAATACTAAGACGTATTGAAATACCTGATAGGAATCAAGTCAAGTAGTAGTAA	4024
Q	1103	tgaataccaagaacgaccaggtatatactatacaacaatgcaaaaatgtaatctgcgcgacaaca	1162
D	4025	TGAAGTCAAAAAAATGATCAAGGAATTAACAATTAATTAATGCAAAATGAATTTTACAGATTAATA	4084
Q	1163	atggtaacgatalcgggttcatcgggttctccacacggttcaacaataatcgttaaacgggtg	1222
D	4085	ATGGGAATGATTAATAGCTTTATATAGATTTTATCATCAGTTTAAATTAATTAATAGCTAAATCAATAG	4144
Q	1223	cttcaaacgtgtacaactcgtcagatcgaaacgttctctcgcacactcgtgggttctgttgg	1282
D	4145	CAATTAATTTGGTAAATTAATAGCAAAATAGAAAGATCTAGTAGCACTTTGGGTTCTCATGGG	4204
Q	1283	agttcatcccggttgtatgaacggttgggtgggaacgtctcgcgtctgaagaatc	1332
D	4205	AATTTATTTCTGTAGATGATGATGGGGGAAAGGCGCACTGTAAATTAATCA	4254

Db 2648 ATGAAAGTAATCATTTAATAGACTTATCTAGGTATGCAATCAAAAATAATATGGTAGTA 2707
 Qy 143 aagtttaacttgcagccgagtaagaaatccagacccgctgtacatcttgaaatcttcca 202
 Db 2708 AAGTAATTTTGGATTCATATAGTAAATAATCAATTTATTATTATTAGAACTAGTA 2767
 Qy 203 aaatcgaagttatccctgaagaatgctacgtatatacactctatgacgaataactctcca 262
 Db 2768 AATTTGAGGTAAATTTTAAATAATGCTATTTGATATATAGTATGTAATAATTTTAGTA 2827
 Qy 263 cctcccttggatccgltacccgaatactccaactccatctctctgaaataatgataca 322
 Db 2828 CTACCTTTGGATTAAGAAATTCCTTAAGTATTTTAAACGATTAATGCTTAATATGAAATATA 2887
 Qy 323 ccatatcaactgcatgataaagaatcttggttggaaagtctctgtaactgagtgtaa 382
 Db 2888 CAATTAATTAATTTGATGGAATAATTAATTCAGATGGAAGTATCACTTAAATTTGGTGAAA 2947
 Qy 383 tcaatctgactctgcagagacactcaggaatacaacaagcgtgtgtatccaatactctc 442
 Db 2948 TAACTGTGACTTTACAGAGTAAGTCTCGAATAATAAACAAGAGTACTTTTAAATACAGTC 3007
 Qy 443 agatgatacaactctgtaactacatcaatcgctggaatctctgttaccatcacaacaatc 502
 Db 3008 AAATGATTAAATATATACATTTATTAACAGATGATTTTGTAACTATCACTAATTAATA 3067
 Qy 503 gtcgaataactccaactctacatccagcgcgtctgcatcgacgaagaacacgactcca 562
 Db 3068 GATTAAATTAATCTTAATAATTTATTAATAATGGAAGATTAATGATCAAAAACCAATTTTCAA 3127
 Qy 563 atctgggttaacatccaacgctcttaataacatcatgttcaactgtagcgtgtgtgaca 622
 Db 3128 ATTAGGTAAATATTCATCTGATATATATATATTTAAATTTAGATGGTTGTACAGATA 3187
 Qy 623 ctacccgctacatctggaatcaactcaatctgttgcacaagaactggaagaaag 682
 Db 3188 CACATAGTAAATATTTGATTAATAATTTTAAATCTTTGATTAAGGAATTAATTAATGAAG 3247
 Qy 683 aaatcaagaagccttgcagacaacacagctccatctgtgtatctctgtaagaactcttggtg 742
 Db 3248 AAATCAAAAGATTTATATATATATCAATCAATCAATTCAGATTTTAAACACTTTTGGGGTG 3307
 Qy 743 actacccctgacagacacaacgcttactacatgctgtaactgtgacatcggaacaatacag 802
 Db 3308 ATTATTTTCAATATGATTAACCAATCACTATATGTTAAATTTATATGATCCAAATTAATATG 3367
 Qy 803 ttgacgtcaacaatgtagtaccggttacaatgttaccgtgaagtcgctgtgttctg 862
 Db 3368 TCGATGTAAATTAATGATGATTTAGAGGTATATGTATCTTAAAGGCGCTAGAGGTAGCG 3427
 Qy 863 ttatgagctacacaactgtaactgtaactctccctgttaccggtgtgataccaatcatcatca 922
 Db 3428 TAAATGACTACAAACCTTTATTTAATTTCAAGTTTGTATATAGGGGCAAAAATTTATTATTA 3487
 Qy 923 agaaatacgcgtctgtgtaacaagaacaatctgttgcacaacaatgatcgtgtatataca 982
 Db 3488 AAAATATATGCTTCTGGAATAATAGATATATTTGTTAGAAATAATGATGCTGATATATATA 3547
 Qy 983 atgttgtatgtaagaacaagaataacgctgtgcttacaatgcttcaagctgtgttag 1042
 Db 3548 ATGTAGTATGTTAAATAATTAAGATATAGTTAGCTAAATGCAATCAGACGACGCGTAG 3607
 Qy 1043 aaagatctgtctgtcctggaatacccggaagcgttggtaactgtctcaggtatgttata 1102
 Db 3608 AAAAATATCTAAGTGCATTAAATAATCTGATGATGGAATAATCTTAAGTAAGTATAGTA 3667
 Qy 1103 tgaataccaagaagaccaggtgtatcacttaacaatgcaaaatgtaactgtgcaggaaca 1162
 Db 3668 TGAAGTCAAAAATGATCAAGAAATTAACAATAATGCAAAAATGATTTTACAGATATATA 3727
 Qy 1163 atgttacaatatacggttcaatccagcttcaacaataatcgctaaactgtgtt 1222

Db 3728 ATGGGAATGATATAGGCTTTATAGATTTTCATGCTTTAATATATAGCTAAACATAGTAG 3787
 Qy 1223 ctccaactggtacacatcgtcagatgaacgcttccctcgcacactggtggtctgttgg 1282
 Db 3788 CAAGTAAATTTGATTAATATAGACAAATAGAAAGATCTTAGTAGGACTTTGGTTCCTCATGGG 3847
 Qy 1283 agtcatcccggttgcagcgttgggtgtaacgcttgcgtgttaa 1326
 Db 3848 AATTTATTTCTGTAGATGATGATGGGGAAGGCACTGTAA 3891

RESULT 12
 AX036248
 LOCUS AX036248 3891 bp DNA linear PAT 16-NOV-2000
 DEFINITION Sequence 27 from Patent Epi041149.
 ACCESSION AX036248
 VERSION AX036248.1 GI:11225866
 KEYWORDS
 SOURCE
 ORGANISM
 synthetic construct.
 artificial sequence.
 1 (bases 1 to 3891)
 REFERENCE
 1 Kink,J.A., Flitca,J.R., Padhye,N.V., Thalley,B.S., Stafford,D.C. and Williams,J.A.
 TITLE
 Vaccine and antitoxin for treatment and prevention of C. Difficile disease
 JOURNAL
 Patent: EP 1041149-A 27 04-Oct-2000;
 OPHIDIAN PHARM INC (US)
 FEATURES
 source
 location/Qualifiers
 1..3891
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 1..3891
 /note="synthetic"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC16480.1"
 /db_xref="GI:11225867"
 /translation="MDFVNFQFNVKDPVNGVDIAVIRKIPNVGOMPVKAFRIHNTKIWI
 IPERTFTNPREGDLNPPPEAKQPVSYSTSTYSTDNKRNILKGVTKLEPRITST
 LGRLMLTISIVGIPWGGSTIDTELKVIDTNCINVTODGSTSESLVLTIGPADP
 IOPEKSPGHVGLNLTNRNGYSTOYIRSPDPTGSESLVDNPLPLGAGKFAFDPA
 VTLAHELHAGRLRGIAINENRVEKVTNAYEEMSGKVEFEELEPTGCHDAFTDS
 LOENEFRLYYNKKFEDIASLTNKAISYGTASLOYKMNVEKYLISEDTSGKESVD
 KIKDKLTKMLTELTYEDNPFVFEKFKYLRKTYLNDPKAFKINIVPKVNTIYGFNL
 RNTNIAANFNQNTPEINMNTFKLKNFTGLIEEYKLLCVRGITTSKRTSLDKRYKAL
 NDLCIKVNMOLFSPSEDNFTNDLNKEELITSDNIEAEBNISLDLQOYLTFENF
 DNEPNSIENLSDIIGLELMPNIEEPNGKVEIDKYTFMFLHRAQEBHGSR
 ALTNSVNEALNPSRVYFFSSDYVKKNKATKAMPLGAWVEOLVYDPTDETSVST
 DKIDITIIPIYIGPALNIGMLKDEDFVALISGAVILLEIPEIAIPVIGFALV
 SYIANKVLTVOITIDNALSKRNEKDEYKIVTNSLAVVLTNOIILIRKKMEALENOA
 EATKALINVOYNOTBEERNININRIDLSKINSIKKAMININPKTINOCSTIAN
 SMTPGVARLEDFDASLKDALLKTYIDNRGLIGVDLKLKRVNNTLSTDIPOLSKI
 VDNORLSTFLEYIKNIINTISILNIRESNHILIDRSASKINIGSKVNFOPIDKNOI
 QLFNLESKIEVILKNAIVYNSMYENFSTFWIRPKYFNSISLNNEETIINLNENNS
 GKVSLNAGEIIMTLODFOELKORVFEVSOMINISDYINMIEVYITNNBLNNSKIY
 INGRILDOKPIPSINGNHISNNINMEKLEKDETHRYIKIKYENLEKLEKNSKMTX
 DNOGNSGILKDFMGDYLOYDKPYMLNLYDNKKYVDVNNVCGIRGMVTKGKRSQWY
 NITYLNSIYRGTKFTIKKYTAGSKNDNIYRNNDRYINIVYVKKERELATNSQAEVK
 ILTSLAETPDVNLSCOVVVKRSKNDGITNKKKMNIDQNNNGMDIGIFGHQPNNTAKLV
 ASNMWRDIESSRILGCSWEFIPVDDGGBRPL"

BASE COUNT 1580 a 392 c 629 g 1290 t
 ORIGIN

Query Match 54.5%; Score 726.4; DB 6; Length 3891;
 Best Local Similarity 72.3%; Pred. No. 1.3e-176;
 Matches 943; Conservative 0; Mismatches 361; Indels 0; Gaps 0;
 Qy 23 tgtctaaccttcctgaatacatcaagaacatcatcaatcccatccctggaaccctgacct 82
 Db 2588 TATCTAATTTACTGAATATATTATTAATTAATCTTATATTGAAATTTAGAT 2647

QY	83	acgataccaataccatcgcacgtcgtctccgcgaagcttccaaatccaacatcggttca	142
Db	2648	ATGAAATGAATCATTTATATAGACTTATCTTAGGTATCATCAAAAATAATATTGGTAACTA	2707
QY	143	aagttaaccttcgcatccgatcgacaaagaatccgatccagctgttcaatcgtgaatcttcca	202
Db	2708	AAGTAATATTTGGATCCCAATGATTAATAAATCAAAATTCAAATTTATTTAGAAAGTAACTA	2767
QY	203	aaatcgaagtataccctgaaagaatgcatatcgatatcaactctatgtcagaaatctctcca	262
Db	2768	AAATTTGAGGTAAATTTTAAAAAAGCGATTTGATATATATGATATGTAAGAAATTTTAACTA	2827
QY	263	ccctcccttcgatccgatcccgaaatcactccaacccaactcctctcgtgaacaaatgaataca	322
Db	2828	CTACCTTTTGGATATAGAAATTCCTTAAGTATTTTACAGGTATTAAGTCTAATATATGAATATA	2887
Db	2888	CAATTAATAAATTTGATGTAATAAATAATTCAGATGAGAAATCATACCTTAATTTATGCTGAA	2947
QY	383	tcactctggaacttcgcgggaacatccagaagaatacaaaagcgtgtgtatctcaatactctc	442
Db	2948	TAACTCTGACCTTTTACAGATATCTCAGAGAAATAAACAAGAGTACTTTTAAATATCACTC	3007
QY	443	agaatgacaacatctcgtgaactacatcaatccgcgtgactcgttccatccaaccaacatc	502
Db	3008	AAATGATTAATATATCGATTAATATTAACAGATGGAATTTTGTAACTATCACTAATAATA	3067
QY	503	gtctgataaactccaaaatctacatcaacagccgctctgtatcgaccagaaacccgatctcca	562
Db	3068	GATTAAATTAATCTTAATAAATTTATATATTAATGAAGATTAATATGATCAAAACCAATTTCAA	3127
QY	563	atctcgggttaacatccacgcgtcttcaataacatcatgttcaactgtgaacgtgtgtgtgca	622
Db	3128	ATTAGGTAAATTAATTCATGATGTAATTAATTAATATGTTTAAATTAATGTTGTATAAGATA	3187
QY	623	ctcacgcctacatcgtgatcaaaatcattcaactcgtctcgacaagaacgtgaacggaaag	682
Db	3188	CACATATATATTTTGGATATAAATATTTTATATCTTTTGTATATGAATTAATATGAATAAG	3247
QY	683	aaatcaagaacgtctacgacaacagttccaatctgtgatctctgaaagaactctcgtgggtg	742
Db	3248	AAATCAAGAATTTATATGATTAATCAATCAAAATCAGATATTTTAAAGACCTTTGGGGCTG	3307
QY	743	actccccgcgagtacgacaacacggtactacatcgtcgaactgtgaagatccgaaacaatacgc	802
Db	3308	ATTATTTTACATATGATTAATAACCATACTAATATGTAATTAATTAATATCACTCAAAATAAATATG	3367
QY	803	ttagacgtcaacaatgttaggtatccgcggtgtacatctgacacgttcgaagaagttccgctgttctg	862
Db	3368	TCGATGTAAATTAATGTRAGTATTAGAGGTTATATGATCTTAAAGAGCCCTTAAGAGTAGCG	3427
QY	863	ttaactacacaaatctactactgaactcttcctcgtacacgtgtgatccaatcatcatcaca	922
Db	3428	TAAATGACTACAACATTTATTTAAATTCAAAGTTTGTATAGGGGACAAATTTATTAATA	3487
QY	923	agaaatacgcgtctgtgaacaagaacaaatcogttcgcacaatgatatcgtgtatacatca	982
Db	3488	AAAAATATGCTCTCGCAAAATAAAGATAAATATTCTTGAGAAATTAATGATGCTGATATATTA	3547
QY	983	atgtgtgaagttaagaacaagaataccgctcgtcgtcacaatgctctccaagctgtgttag	1042
Db	3548	ATTGATGATTTTAAAAATAAAGAAATTAATAGGTATATGCTATATCATCACAGCGAGGCGTAG	3607
QY	1043	aaaagaatcttgcgtcgtcgtgaatcccgagcgtgtgtaatctgtcgaagtgaagtga	1102
Db	3608	AAAAAATTAATTAAGTATTGAAATATCTTGATGTAGAAATCTTAAGTCAAGTAGTAGTA	3667
QY	1103	tgaatccaagaacgcacaggtgatcatcaactacaacatgycanaatgtaatctgcagcgaca	1162
Db	3668	TGAAGTCAAAAAATGATCAAGGAATTAACAATAATTAATGCAAAATTAATTAACAAGATATA	3727
QY	1163	atgtgtacgatatcgtttcatcgtgttccaccagttcaacaatatcgtctaactgtgtg	1222

Db	3728	ATGGGAATGATATGCGTTTATAGATTTCATCTCACTTTAATAATATTTACTTAACATAGAG	3787
Oy	1223	ctccaactggatcaatcgtcagatcgaaacgcttcctcgcgaactcgtgggtctcctgg	1282
Db	3788	CAAGTAATTTGGTATTAATAGACAATAAGAAATCATAGACACTTTGGGTGCTCATGGG	3847
Oy	1283	agttcatcccggttgatgaacggttggggctgaagcgcgctgtaa	1326
Db	3848	AATTTATTTCTGTAGATGATGATGGGGAAGGACCACTGTAA	3891

RESULT	13
CENOTOXA	
LOCUS	CBNT0XA 4067 bp DNA linear BCT 24-JUN-1994
DEFINITION	C.botulinum gene for infant neurotoxin type A.
ACCESSION	X73423
VERSION	X73423.1 GI:507070
KEYWORDS	botulinum neurotoxin; botulinum neurotoxin type A.
SOURCE	Clostridium botulinum.
ORGANISM	Clostridium botulinum. Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
REFERENCE	1 (bases 1 to 4067) Williams, A., East, A.K., Lawson, P.A. and Collins, M.D.
AUTHORS	Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A associated with infant botulism: comparison with other clostridial neurotoxins
TITLE	Res. Microbiol. 144 (7), 547-556 (1993)
JOURNAL	2 (bases 1 to 4067) Williams MCG, A.
MEDLINE	Direct Submission
REFERENCE	Submitted (17-JUN-1993) A. Williams MCG, AFRC Institute of Food Research, Dept of Microbiology, Reading Laboratory, Earley Gate, Whiteknights Road, Reading, RG6 2EF, UK
JOURNAL	Location/Qualifiers
FEATURES	1. 4067
SOURCE	/organism="Clostridium botulinum" /strain="Kyoto-F" /db_xref="taxon:1491" 83. 88 97. 3987 /codon_start=1 /transl_table=11 /product="botulinum neurotoxin type A" /protein_id="CA51824.1" /db_xref="GI:507071" /db_xref="SPRMBL:045894" /translation="MPVYNQGFNRKQFNVGVDIAVYIKIPNAGQMPKAKIKHKIMW IPEKDTNPDEGLNDPPEKQKQVPVSYSTYSTONENKNTLKVTKLPERIYSTD LGRRLTSIVRIGLIPFWGSGTIDELKIDINCINWIPDPSYSEBELNVIIGSADID IOEFCSEFGHDVNLTRNGYSTQYIFRSPDFTFGESELEVDNPEELIAGKFAITDPA VTLAHELIAHEDRLYGTAINPNRVEKNTANAYMSGLESEFEELPRFEGHDAFISD LOEKEPFLYYNKRKFDQAISLTKAKSLIGTATASQWAKNPKKEYLLSEPTSGESVD KLEKDKLTKMLTELYTEDNENFVKVYINRKYTLNFDKAVRINIVPEPNTIKGFNL KCAALSTNFGQNTLEINSRNFTRLKNTGTEFEYKLLCAGVGIIPKYSIDEGYNKAL NDLICIKVNMNDLEFSPEDNEFTNDLVEEITADTNEAEENISDLIOOYLYTFPPE DNENPESISENLSIDIGOLEPMNIEPRFNGKKELEDKTDMFHYLAQOEHEHDSRI ILNRSAEALIKPVNVAATPFSSQYKVKINKAVAEFMLEAEELVYOFETNEVTVM DKIADITIIPTPIGALNIGMLSKGEVEALIFTGVVAMELPIPEVALYEGFALIV SYLANKLVVQITNNALSKRNEKMEYKTKVNNLAKVNTQIDLIEKKMKALENOA EAKRKAIVNOYNTTEEEKNNINENIDLSLWESINSAMININRKLDCQSVYLAN SMPIAVKRLKDPASVADVLAKYIDNRGTLVQVDRKDEVAENTLSADIPQLSKRY VDMKLIKSTFEYEVKNIIVNTSILSIYKDKDLDLSRYGAKINIGRWYDSDIKNDID KLINLESSTIEVILKNAIVYNSMTENSTSPWIKIPKPYFSKININNEYTIIINCENNS GKWSLVNGETIWTLODNKQITQVREKISQMVNISDYNRMIEVYTTNNRLTKSKLY INGLIDQKRIISNLGNHASNKIMFKLDGCRDPRRYIMYIENFLDEKLEKEIKDLY DYSQNSGILKDPFNGNIQYDKRPYMYMLPDPNKYDVNNIGIRGYMLKRGSGSVTTT NIYSNLSLEGTEKEIIRKYVSGNEDNIVRNNDRYINVVAKREYRLATNAGSGVGEK ILSLAETIPDYGNIISOVYVMSKDDOGIRNCKMNLQDNNGNDIGFIGPHLYDNTAKLV ASNRYNQGVRKASRSTFGCCSNDFIIPVDGMESSLS"

misc_feature
1384. .1386

	misc_feature	/note="C residue (S-S bonding)"	
	misc_feature	1440/1441	
	misc_feature	/note="possible cleavage position L-H"	
	misc_feature	1456..1458	
		/note="C residue (S-S bonding)"	
	BASE COUNT	1640 a 407 c 664 g 1356 t	
	ORIGIN		
	Query Match	49.7%; Score 662; DB 1; Length 4067;	
	Best Local Similarity	69.1%; Pred. No. 5..5e-160;	
	Matches 905; Conservative 0; Mismatches 405; Indels 0; Gaps 0;		
Oy	23	tgcttaaccttcctgaataatcacaagaatcaataatcctccatcctgaacctgcgt	82
Db	2664	TATCTACATTTTACTCAATTAATTAATTAAGAAATTTGTTAAATCCCTTAATTAAGATTAATGAT	2743
Oy	83	acgaatccaatcaactgaatcgacccgtgcctgcctacgcctcccaaaatcaacatcoggttca	142
Db	2744	ATTAATAAGATGATTTTAATAGATTTTATCTAGGTATGACCAAAATAATATTTGGGGATA	2803
Oy	143	aagttaaactcgatccgatcgacaagaatcagatccagctgltcaatctgaaatcttcca	202
Db	2804	GAGTATATTTATATTCATCAATAGATTAATAAATCAATTAATTAATTAATTTAGAAAGTAGTA	2863
Oy	203	aatcgaaagtctcctgaagaatgctatcgttatcaaccttatgtaacgaacttcca	262
Db	2864	CAATTGAGGTAAATTTTAAAAAATGCTATTGTATTAATTAATGTAATGTAAGAAATTTTAGTA	2923
Oy	263	cctctctctgtaatccgtatccgaaatctcaactccatcctctcgacaatgaatata	322
Db	2924	CTAGGTTTTGGATTAATAAATTCCTAGTATTTTAGCAAGATTAATCTAATTAATGATTA	2983
Oy	323	ccatcatcaactcgatcgtgaaacaaactctggtltggaagtatcctgaactcogltgaa	382
Db	2984	CAATATATAATTTGTATAGAAATAATTTTCAGATGGAATAATATACCTATTATTTAGTGAA	3043
Oy	383	tcatctgaactctgacagacactcaggaatacaaaagcttggatccaatctatctc	442
Db	3044	TAAATCTGCACTTTGCAGGATTAATTAAGCAAAACATTAAGAGAGATTTTAAATTAACATG	3103
Oy	443	agatgatcaacaatctctgatactatcaatcaatcgcgtgtaatctcgtttacatcaaccaaacat	502
Db	3104	AAATGTTTAATTAATTCAGATTAATTAACAGATGATGATTTTGTAACTATACATTAATTA	3163
Oy	503	gtctgaataactcccaaaatctacatcaacacgcgcgtctgtatcgaccagaacogatcca	562
Db	3164	GACTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	3223
Oy	563	actctggtaacatccaacgcttcttaataacatcatgltcaaacctgacggtgttgcgtgaca	622
Db	3224	ATTTAGGTAAATTTTCATGCTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	3283
Oy	623	ctcacgcgtacatctgtaataaatacttcaatctgltcgacaaagaactgaaacgaaaga	682
Db	3284	CACGTAGATACATCAATGAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTA	3343
Oy	683	aaataaagaactctgaacgaacacgcgtccaattctggtatccttgaaagactctggtgtg	742
Db	3344	AAATCAAAATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	3403
Oy	743	actaccctgagtagcacaacacgtaactacatgctgaatctgtagcatctcgacaacataag	802
Db	3404	ATTATTTTACAAATATGATTAACATACATCAATATATGTTAAATTTATTTGATCCAAATTAAT	3463
Oy	803	ttgacogtcaacaatgtaggtatccgcggtgtacatgtaactgaaaggltccgcgtggtcttg	862
Db	3464	TCGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	3523
Oy	863	ttaatgataccaactatcactcgaactctcctctgtaacogtgggtacccaattcatcata	922
Db	3524	TAGTGCATCAACAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	3583

BASE COUNT	ORIGIN
397 a	342 c
207 g	353 t

OY	1038	tgtagaagaagatcctgctgcctcgaaatcccggaagtgttaa---tcgtccaagt	1094
Db	1041	ACCGGAAAAAATCATCAAACTATCCGTACTTCTTAACCTCTCTGGTGACAGT	1100
OY	1095	agttgtaatgaataaccagaacgaccaggatcatcaactaaatgcaaatgaattcga	1154
Db	1101	CATCGTTATTGACACTC-----GATCGGTAAACAACCTGCATGAACCTTCCA	1145
OY	1155	ggacaacaatggttaacgatattcggttcttggtttccaccagctccaatatatgctaa	1214
Db	1146	GAAACAACAACGGGTGATACATGGTCGTGGGTTCACCTTAACAAAC-----	1194
OY	1215	actggttgttccaactggttacatcgcgttagatcgcaagcttcctcgcacgtcgtgty	1274
Db	1195	-CTGGTGTCTTCTTCTGTGTACTAACACAACATCCGTAATAAACACTTCTTTACGGTTG	1253
OY	1275	cctctgggaagttcacctccggttgatgaacggttgagggtgaa	1314
Db	1254	CTTGTGGTCTTTCAATCTCTAAGAACAACGGTTGGCAGGAA	1293
RESULT	15		
LOCUS	A58946	1313 bp	DNA linear PAT 06-MAR-1998
DEFINITION	Sequence 6 from Patent WO9641881.		
ACCESSION	A58946		
VERSION	A58946.1 GI:3714418		
KEYWORDS			
SOURCE	unidentified.		
ORGANISM	unclassified.		
REFERENCE	1 (bases 1 to 1313)		
AUTHORS	Elmore,M.J., Mauchline,M.L., Minton, Nigel,P. and Pasechnik,V.A.		
TITLE	TYPE F BOTULINUM TOXIN AND USE THEREOF		
JOURNAL	PATENT: WO 96/41881-A 6 27-DEC-1996;		
COMMENT	MICROBIOLOGICAL RES AUTHORITY (GB)		
FEATURES	Other publication AU 6012196 970109.		
source	location/Qualifiers		
	1..1313		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	401 a 344 c 210 g 358 t		
ORIGIN			
Query Match	31.7% Score 421.6; DB 6; Length 1313;		
Best Local Similarity	62.2%; Pred. No.5.8e-98;		
Matches 809; Conservative	0; Mismatches 434; Indels 57; Gaps		
OY	30	cttcaatgaatcacaaagaacatcatcaatccctccalcctcyaacctgcyctacgaatc	89
Db	44	CTTCAACAACCTGTACAAAATAAATCAAAGACAACTCTATCTCGGACATGCCGTACGAA	103
OY	90	caatcacctgatcgacctgctgcgtctacgcttccaaaatcaaacatcggttctaagttaa	149
Db	104	CAACAATTTCAATCGACATCTCGGCTATGGTTCTTAACATCTCTATCAACGGGACGTGA	163
OY	150	cttgatccggaatcgcaagaatlcagatcagctgfttcaatctcgyaaacttccaaatcga	209
Db	164	CATCTACTCTACTAACCCGCAACCACTGGGTATCTACTCTTTAAACCCTCTGAAGTAA	223
OY	210	agttacttcgaagaatgctatcgtlatatacaactctatgtaagaaacttctccacctct	269
Db	224	CATCGCTCAGAAACAACGACATCATCTACACAGCGTGGTACAGAACTTCTATCTCTT	283
OY	270	cttgatccggtatcccgaataacttaactcaatctctcttgaaacatgataacacatac	329
Db	284	CTGGGTGTGTATCCCGGAATACTTCAACAAAGTTAACTGAACCAACGAATACATCAT	343
OY	330	caactgcat----ggaacaaatctcgygttggaagaatctctgaaactacggtlgaatcat	386
Db	344	CGAGTCATCCGTAAACAACAACCTGGTTGGAAAATCTCTCGAATCTAAACAANAATCAT	403

OY 387 ctggaactctgcaggaacactcaagaaatcaaacagcgtgtgtatccaataactctcaagt 446
|||||
Db 404 CTGGACTCTGCGAGGACACTGCTGTATACACAGAAACGTTTCAACTACATCAGAT 463
OY 447 gatcaacatctctgaactacatcatcgcgtgaatcttcgtttaccatccaacatcgtct 506
|||||
Db 464 GATCTCTATCTCTGACTACATTAATGAATGATCTTCGTTACTATCACTAACACCGTCT 523
OY 507 gaataactccaatctcaatcatcagcgcgtgtgtatcgaccagaaacggatctccaatct 566
| |||||
Db 524 GGGTACTCTGCTATCTATCTATCAACGGTACCTGATCGATGAATAATCTATCTTACCT 583
OY 567 ggttaacatccacgccttcttaatacatcatgttcaaacctgacggttgcgtgacactca 626
|||||
Db 584 GGGTGACATCCAGCTTTCGACACATCCGTTCAAAATGTTGGTTCACAGACAC--- 640
OY 627 ccgctaacatctggaatcaaatcaatctcgttgcacaaagaaactgaaacgaaagaat 686
|| |||
Db 641 GCGTTACGTTGGTATCCGTTACTCAAAAGTTTTCGACACTGAACCTGGGTAAACTGAAT 700
OY 687 caaagacctgtacgacaacacagtcacatctcgttatccctgaagaactctgggtgacta 746
|| |||
Db 701 CGAAACTCTGTACTGTGACGAAACCGACCGCTTATCTGANAAGACTTTCGGGTACTTA 760
OY 747 cctgagtaacgacaacacgtaactacatcgtgaatctgtacgaaccgaacaatacgtga 806
|||||
Db 761 CTTGCTGTACACAAACGTTACTACTGTAACCTGCTCGGACCTGACAAATCTATCAG 820
OY 807 cgtcaacaatgtaagtaaccgcggttaacatgtaacctgaaaggtccgcgtgtgtctgtat 866
|| |||
Db 821 TCAGAACTCT-----AACTTCCTGAACATCAACACGACGCGGTGTTTATCA 868
OY 867 gactaccaacatctacactcgtgaactctccctgtacggtgtgacaaatcatcatcaagaa 926
|| |||
Db 869 GAAACCTAATATCTCTCTAACACTCGTCTGTACACTGCTGTGAAGTATATCATCCGTA 928
OY 927 ataagcgtct-----gtaacaaggaacaatacgttgcgaacaacatgatacgtgata 977
|| |||
Db 929 AAACGTTTCTACTGACATCTGACACTGACACTGACACTGTAATAAAACGACCTGGCTTA 988
OY 978 catcaatltgtgaatgaacaagaataacgctcgtgtacacaaatgctctcaggtg 1037
|||||
Db 989 CATCAACGTTGTTGACCGGTGACGTTGAATACCGTCTGTACGCTGACATCTTATCGCTAA 1048
OY 1038 tgtagaagaagatctgtcgtctgtgaaatcccggaacgttggtaa---tctgtcgaagt 1094
|||||
Db 1049 ACCGGAATAATCATCAAACTGATCCGTACTCTTAACACTCTAACAACTCTGGGTGAGAT 1108
OY 1095 agtltgaatgaatccaagaacgacgaaggtatcaactaacaatgcaaaatgaatctgca 1154
|| |||
Db 1109 CATCGTTATGGACTC-----GATCGGTAACTGACACTGACATGAACTTCCA 1153
OY 1155 ggaacaagaatgtgaagatcgtgttcaatcggtttccacacagttccaacaataatcgctaa 1214
|||||
Db 1154 GAAACAACAACGGGTGTAACATGGTCTGCTGGGTTTCCACTTAACAAAC----- 1202
OY 1215 actggttgcctcaactgtgtacatcgtlcaagtcgaagttcctcctcgaactcgtggtg 1274
|||||
Db 1203 -CTGGTGTCTTCTTCTGTGTACTACACAAACATCCGTAAAAACACTTCTTCTAACGGTTG 1261
OY 1275 ctcttggaagttcatccggttgatgaacggttggggtgaa 1314
|| |||
Db 1262 CTTCTGTCTTCTCATCTCTAAAGAACACGCTTGCGAGGAA 1301

Search completed: September 16, 2002, 21:29:41
Job time: 15573 sec

THIS PAGE BLANK (USPTO)

Wed Sep 18 09:34:26 2002

us-09-611-419a-1.rng

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 18:07:58 : Search time 631.73 Seconds
(without alignments)
3620.107 Million cell updates/sec

Title: US-09-611-419A-1

Perfect score: 1332

Sequence: 1 gaatcgaacagatgcgtct.....aacgtccgtcgaagaattc 1332

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N.Geneseq_032802:*

1: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT:*
2: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT:*
3: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT:*
4: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT:*
5: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT:*
6: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT:*
7: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT:*
8: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT:*
9: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT:*
10: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT:*
11: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT:*
12: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT:*
13: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT:*
14: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT:*
15: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT:*
16: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT:*
17: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT:*
18: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT:*
19: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT:*
20: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT:*
21: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT:*
22: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1332	100.0	22	AAA54482	Botulinum toxin hea
2	1311	98.4	17	AAT29245	Type A neurotoxin
3	1311	98.4	1330	AAV30571	Clostridium botuli
4	1311	98.4	21	AAZ87212	DNA encoding synth
5	1311	98.4	1351	AAV30576	Clostridium botuli
6	1311	98.4	1402	AAT29245	Type A neurotoxin
7	1311	98.4	1402	AAV30572	Clostridium botuli
8	1310.4	98.4	1323	AAA54483	Botulinum toxin hea
9	1308.2	98.2	22	AAA54484	Botulinum toxin hea

10	1256.6	94.3	1332	22	AAA54588	Sequence encoding
11	727.6	54.6	1546	19	AAV30575	Clostridium botuli
12	727.6	54.6	4835	21	AAC64582	BoNT/A neurotoxin
13	726.4	54.5	1317	21	AAZ87220	DNA encoding nativ
14	726.4	54.5	2532	21	AAZ87218	DNA encoding nativ
15	726.4	54.5	3891	17	AAT29244	C. botulinum type
16	694	52.1	702	21	AAZ87221	DNA encoding BoNTA
17	619	46.5	621	21	AAZ87222	DNA encoding BoNTA
18	421.6	31.7	1313	18	AAT48101	Immunogenic type F
19	421.6	31.7	1314	22	AAA54499	Botulinum toxin hea
20	401.6	30.2	1317	21	AAZ87216	DNA encoding synth
21	399.4	30.0	1317	22	AAA54490	Botulinum toxin hea
22	371.8	27.9	1917	21	AAZ87219	DNA encoding nativ
23	289.4	21.7	1347	21	AAZ87217	DNA encoding synth
24	289.4	21.7	1368	22	AAA54491	Botulinum toxin hea
25	273.2	20.5	1278	22	AAA54489	Botulinum toxin hea
26	264.8	19.9	1347	22	AAA54589	Sequence encoding
27	263	19.7	1341	21	AAZ87213	DNA encoding synth
28	263	19.7	1400	22	AAA54485	Botulinum toxin hea
29	262.4	19.7	1400	22	AAA54488	DNA encoding synth
30	262.4	19.7	1400	22	AAA54488	Botulinum toxin hea
31	227.2	17.1	1293	18	AAT48100	Immunogenic type F
32	202.8	15.2	1374	22	AAA54487	Botulinum toxin hea
33	193.4	14.5	1472	19	AAV30585	Clostridium botuli
34	193.4	14.5	4017	22	AAV58862	C botulinum BoNT/E
35	190.2	14.3	1463	19	AAV30584	Clostridium botuli
36	180.8	13.6	1460	19	AAV30593	Clostridium botuli
37	168.8	12.7	1371	21	AAZ87214	DNA encoding synth
38	168.8	12.7	1371	22	AAA54486	Botulinum toxin hea
39	166.4	12.5	1359	12	AAO12121	Synthetic tetanus
40	166.4	12.5	3712	16	AAO97490	Plasmid pHTRA1. N
41	166.4	12.5	3754	15	AAO57880	Intermediate plasm
42	166.4	12.5	3754	16	AAO85420	Plasmid pTECH1. S
43	166.4	12.5	3769	15	AAO57881	Intermediate plasm
44	166.4	12.5	3769	16	AAO97492	Plasmid pTECH2. N
45	166.4	12.5	4366	16	AAO85424	Plasmid pTECH3-P28

ALIGNMENTS

RESULT 1	
AAA54482	
ID AAA54482 standard; DNA; 1332 BP.	
XX	
AC AAA54482:	
XX	
DT 11-APR-2001 (first entry)	
XX	
DE Botulinum toxin heavy chain C-terminal coding sequence (serotype A).	
KW Botulinum toxin; neurotoxin; heavy chain; recombinant expression;	
KW recombinant vector; antigen; immune response; vaccine; bacterium;	
KW infection; ds.	
XX	
OS Synthetic.	
OS Clostridium botulinum.	
XX	
FH key	Location/Qualifiers
FT CDS	13..1326
FT	/*tag= a
FT	/product= H_C peptide fragment
XX	
PN WO200067700-A2.	
XX	
PD 16-NOV-2000.	
XX	
PF 12-MAY-2000; 2000WO-US12890.	
XX	
PR 12-MAY-1999; 99US-0133865.	
PR 12-MAY-1999; 99US-0133866.	
PR 12-MAY-1999; 99US-0133867.	
PR 12-MAY-1999; 99US-0133868.	

Sept for
11/1/02
Sequence
Page 1
Spec

XX OS Synthetic.
 XX FH Key
 FT CDS
 FT 1.1317
 XX Location/Qualifiers
 XX /tag= a
 XX W09612802-A1.
 XX PN
 XX PD 02-MAY-1996.
 XX PF 23-OCT-1995; 95WO-US13737.
 XX PR 07-JUN-1995; 95US-0480604.
 XX PR 24-OCT-1994; 94US-0329154.
 XX PR 16-MAR-1995; 95US-0405496.
 XX PR 14-APR-1995; 95US-0422711.
 XX PA (OPHI-) OPHIDIAN PHARM INC.
 XX PI Ficca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
 XX PI Williams JA;
 XX DR WPI. 1996-230603/23.
 XX DR P-PSDB; AAR95008.
 PT Fusion proteins comprising non-toxin protein and part of toxin
 PT useful to form anti-toxins against Clostridium botulinum type A, and
 PT C. difficile type toxins, and to treat C. difficile intoxication,
 PT partic. diarrhoea
 XX
 XX PS Example 22: Page 336-38; 434pp; English.
 CC A synthetic gene (AAR95008) codes for the heavy chain C fragment
 CC (AAR95008) of Clostridium botulinum type A neurotoxin (see also
 CC AAR95010). Codon usage allowing efficient gene expression in Escherichia
 CC coli was utilised. The gene in vector pAlterBot was used to make
 CC expression constructs in which fragments of C. difficile toxin A
 CC repeat domains were expressed as genetic fusions with the C.
 CC botulin C fragment and expressed in E. coli.
 XX
 XX SO Sequence 1330 BP, 400 A; 339 C; 246 G; 345 T; 0 other;

Query Match 98.4%; Score 1311; DB 17; Length 1330;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 cgtctgctgtctactcttcaatgataatcaagaacatcatcaatcaatctcctgaac 75
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 7 cgtctgctgtctactcttcaatgataatcaagaacatcatcaatcaatctcctgaac 66
 QY 76 ctgctgctgaatccaatctgatactgacgtctcgttaagcttcaaatcaatc 135
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 67 ctgctgctgaatccaatctgatactgacgtctcgttaagcttcaaatcaatc 126
 QY 136 gcttcaagaatctgatactgacgtcgaagaacatcagatcagctgttcaatcggaa 195
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 127 gcttcaagaatctgatactgacgtcgaagaacatcagatcagctgttcaatcggaa 186
 QY 136 tcttcaaaaatcgaagtatctctgaagaatgctatcgtatataactctatgtaagaana 255
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 187 tcttcaaaaatcgaagtatctctgaagaatgctatcgtatataactctatgtaagaana 246
 QY 256 tcttcaactcttctgtgatactgatactcgaagaataactcaatctccttgaacaat 315
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 247 tcttcaactcttctgtgatactgatactcgaagaataactcaatctccttgaacaat 306
 QY 316 gaatacaccatcatcaactgatactgatactcgaagaataactcgttgaagaatctcgaactac 375
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 307 gaatacaccatcatcaactgatactgatactcgaagaataactcgttgaagaatctcgaactac 366
 QY 376 ggtgaataatcgtgactgtgacgacactcgaagaataatcaaacagcgtgttattatcaaa 435

DB 367 ggtgaataatcgtgactgtgacgacactcgaagaataatcaaacagcgtgttattatcaaa 426
 QY 436 tactctcaagatgatacaatctctgatactatcaatccgctgtgattctgttccatcacc 495
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 427 tactctcaagatgatacaatctctgatactatcaatccgctgtgattctgttccatcacc 486
 QY 496 aacaatcgtctgaataatcaatccaaatctcatcaatcagcgcgtctgatactcgaagaacg 555
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 487 aacaatcgtctgaataatcaatccaaatctcatcaatcagcgcgtctgatactcgaagaacg 546
 QY 556 atctccaatctggtgataatcaatccagcttcttaataatcaatcgttcaactggaagctgt 615
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 547 atctccaatctggtgataatcaatccagcttcttaataatcaatcgttcaactggaagctgt 606
 QY 616 cgtgacactcaccgctgatactgatactcaatctcgttcaactggaagctgtgac 675
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 607 cgtgacactcaccgctgatactgatactcaatctcgttcaactggaagctgtgac 666
 QY 676 gaaaagaataatcaaatcagctgatactcgaagaacagcttcaatctgatactcgaagaac 735
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 667 gaaaagaataatcaaatcagctgatactcgaagaacagcttcaatctgatactcgaagaac 726
 QY 736 tgggtgactgactcgtgatactcgaagaacagcttcaatctgatactcgaagaac 795
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 727 tgggtgactgactcgtgatactcgaagaacagcttcaatctgatactcgaagaac 786
 QY 796 aataagctgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 855
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 787 aataagctgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 846
 QY 856 ggtctgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 915
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 847 ggtctgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 906
 QY 916 atcatcaagaataatcagctgatactcgaagaacagcttcaatctgatactcgaagaac 975
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 907 atcatcaagaataatcagctgatactcgaagaacagcttcaatctgatactcgaagaac 966
 QY 976 tcatcaatggtgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 1035
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 967 tcatcaatggtgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 1026
 QY 1036 ggtgataagaatcgtctgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 1095
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1027 ggtgataagaatcgtctgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 1086
 QY 1096 gttgataagaatcgtctgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 1155
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1087 gttgataagaatcgtctgatactcgaagaacagcttcaatctcgttcaatctgatactcgaagaac 1146
 QY 1156 gaaacaatggtgatactcgttcaatctcgttcaatctcgttcaatctcgttcaatctcgttcaat 1215
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1147 gaaacaatggtgatactcgttcaatctcgttcaatctcgttcaatctcgttcaatctcgttcaat 1206
 QY 1216 ctggtgcttccacacgtgatactcgttcaatctcgttcaatctcgttcaatctcgttcaatctcgttgc 1275
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1207 ctggtgcttccacacgtgatactcgttcaatctcgttcaatctcgttcaatctcgttcaatctcgttgc 1266
 QY 1276 tcttggagatcattcccggttgaatgagcttgggttgaatgagcttgggttgaatgagcttgggttgaat 1326
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 1267 tcttggagatcattcccggttgaatgagcttgggttgaatgagcttgggttgaatgagcttgggttgaat 1317

RESULT 3
 AAV30571
 ID AAV30571 standard; DNA; 1330 BP.
 XX
 XX AAV30571;
 AC
 XX 07-DEC-1998 (first entry)
 DT
 XX Clostridium botulinum toxin A fragment C gene in pAlterBot.
 DE

```

XX AntiToxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulinum; ds.
XX Clostridium botulinum serotype A.
OS
FH Key Location/Qualifiers
FT CDS 1..1317
FT misc_difference 1..6 /tag= a
FT /tag= b
FT /note= "PALIER vector-derived nucleotides
ET (encode Met-Ala)"
PN WO9808540-A1.
XX
XX 05-MAR-1998.
XX
XX 28-AUG-1997; 97WO-US15394.
XX
XX 28-AUG-1996; 96US-0704159.
XX
XX (OPHI-) OPHIDIAN PHARM INC.
XX
XX Thalley BS, Williams JA.
XX
XX WPI; 1998-230234/20.
XX
XX P-PSDB; AAM68389.
XX
XX Host cell containing recombinant expression vector encoding
XX Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX
XX Example 22; Page 262-263; 428pp; English.
XX
XX This is the DNA sequence of the Clostridium botulinum serotype A
XX toxin C-fragment gene contained in plasmid palterBot. Recombinant
XX C-fragment proteins have been produced in Escherichia coli as
XX fusion proteins with either maltose binding protein or
XX Clostridium difficile type A toxin (see AAM68387). The invention
XX relates to recombinant proteins derived from C. botulinum toxins.
XX Methods are provided which allow for the isolation of soluble
XX recombinant proteins free of significant endotoxin contamination.
XX Preferred hosts for production of recombinant proteins are E. coli,
XX insect cells and yeast cells. The recombinant toxin proteins are
XX used as immunogens for the production of vaccines and antitoxins
XX that are useful in the treatment of humans and animals at risk of
XX intoxication with clostridial toxin.
XX
XX Sequence 1330 BP; 400 A; 339 C; 246 G; 345 T; 0 other;
SQ

```

```

Query Match 98.4%; Score 1311; DB 19; Length 1330;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 cgtctgctgtctaaccttactgaataacataagaacatcatcaatactcctcgaac 75
DB 7 cgtctgctgtctaaccttactgaataacataagaacatcatcaatactcctcgaac 66
QY 76 ctgctgctgaatcaatcaatcctgctgacctgtctcgtcaagcttccaaaatcaatc 135
DB 67 ctgctgctgaatcaatcaatcctgctgacctgtctcgtcaagcttccaaaatcaatc 126
QY 136 ggtccaaagttaactcgatcgatcgacaaagaatcagatccagctggtcgaatcgaa 195
DB 127 ggtccaaagttaactcgatcgatcgacaaagaatcagatccagctggtcgaatcgaa 186
QY 196 tcttccaaaatcgaaagttaactcgatcgaaagtctatcgatatacaactctatgtaacga 255
DB 187 tcttccaaaatcgaaagttaactcgatcgaaagtctatcgatatacaactctatgtaacga 246
QY 256 ttcttccaaactctctggtacgcgtatccggaataacttcaactccatctctcgaacat 315

```

```

DB 247 ttcttccaaactctctggtacgcgtatccggaataacttcaactccatctctcgaacat 306
QY 316 gaatacaccaatcaatcaatcgtatcgatcgaaacaaatctcgttggaaagtatctcgaactac 375
DB 307 gaatacaccaatcaatcaatcgtatcgatcgaaacaaatctcgttggaaagtatctcgaactac 366
QY 376 ggtgaataatcaatcgtatcgatcgatcgaaacaaatctcgttggaaagtatctcgaactac 435
DB 367 ggtgaataatcaatcgtatcgatcgatcgaaacaaatctcgttggaaagtatctcgaactac 426
QY 436 tacttccaaagttaactcaatctctggtatcaataatcgtcgtggttcttaccatcac 495
DB 427 tacttccaaagttaactcaatctctggtatcaataatcgtcgtggttcttaccatcac 486
QY 496 aacaatcgcttgaataatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 555
DB 487 aacaatcgcttgaataatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 546
QY 556 atctccaatctggtgaataatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 615
DB 547 atctccaatctggtgaataatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 606
QY 616 cgtgaacatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 675
DB 607 cgtgaacatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 666
QY 676 gaaagaagaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 735
DB 667 gaaagaagaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 726
QY 736 tgggtgtaactcctgtaactcgaataacgcgtactcaatcgtatcgtatcgtatcgtatcgtatc 795
DB 727 tgggtgtaactcctgtaactcgaataacgcgtactcaatcgtatcgtatcgtatcgtatcgtatc 786
QY 796 aatatcgttgaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 855
DB 787 aatatcgttgaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 846
QY 856 ggtctgttgaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 915
DB 847 ggtctgttgaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 906
QY 916 atcatcaagaataacgcgttctgtaactcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 975
DB 907 atcatcaagaataacgcgttctgtaactcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 966
QY 976 tacatcaatctgtgtaactcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1035
DB 967 tacatcaatctgtgtaactcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1026
QY 1036 ggttgaagaagaatctgtctgtctgtgaataatcccggaatctgtgaatctgtctgaatga 1095
DB 1027 ggttgaagaagaatctgtctgtctgtgaataatcccggaatctgtgaatctgtctgaatga 1086
QY 1096 gttgttaatgaataatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1155
DB 1087 gttgttaatgaataatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1146
QY 1156 gacaacaatgtaagaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1215
DB 1147 gacaacaatgtaagaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1206
QY 1216 ctggttgccttcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1275
DB 1207 ctggttgccttcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 1266
QY 1276 tcttgggaatctatcccggttgaatcgaatcgttgggtggaacgctcgtcgtgtaa 1326
DB 1267 tcttgggaatctatcccggttgaatcgaatcgttgggtggaacgctcgtcgtgtaa 1317

```

RESULT 4

AA87212
 ID AA87212 standard; DNA; 1338 BP.
 XX
 AC AA87212;
 XX
 DT 08-MAY-2000 (first entry)
 XX
 DE DNA encoding synthetic BONT serotype A (BONTA) Hc fragment.
 XX
 KM Botulinum neurotoxin; heavy chain; BONT; serotype A;
 KM C-terminal fragment; Venezuelan equine encephalitis virus replicon;
 KM VEE; botulism; vaccine; diagnosis; drug screening; ds.
 XX
 OS Clostridium botulinum.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 9..1325
 FT /tag= a
 FT /product= "Synthetic botulinum neurotoxin serotype A
 (BONTA) heavy chain C-terminal fragment (Hc)"
 XX
 PN WO200002524-A2.
 XX
 PD 20-JAN-2000.
 XX
 PF 09-JUL-1999; 99WO-US15570.
 XX
 PR 10-JUL-1998; 98US-0092416.
 PR 12-MAY-1999; 99US-0133870.
 XX
 PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
 XX
 PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
 XX
 DR WPI: 2000-160827/14.
 DR P-PSDB: AAT77134.
 XX
 PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
 PT toxin serotypes A-G, is used for inducing an immune response against
 PT botulinum -
 XX
 PS Disclosure: Page 54; 54pp; English.
 XX
 CC The invention relates to novel vaccines that induce a protective immune
 CC response against botulinum neurotoxin (BONT) serotypes A, B, C, D, E, F
 CC and G (BONTA-BONTG). The vaccine of the invention is novel recombinant
 CC DNA construct comprising a vector, and at least one nucleic acid
 CC fragment comprising a C-terminal heavy chain fragment (Hc) from BONT
 CC serotypes A-G. In preferred embodiments of the invention, the vector is
 CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
 CC this vector results in the production of large amounts of a protein
 CC encoded by a sequence cloned into the replicon. The constructs are used
 CC to produce vaccines against botulism. The proteins can also be used as
 CC diagnostic tools for the diagnosis of botulism. The transformed host
 CC cells can be used to analyse the effectiveness of drugs and agents which
 CC inhibit toxin effects. The vaccine currently used against botulism is
 CC dangerous and expensive to produce, and contains formalin, which is very
 CC painful for the recipient. Also, the vaccine is incomplete, in that only
 CC 5 of the 7 serotypes are represented in the formulation. The novel
 CC vaccine of overcomes these problems, as it is easily purified, and
 CC available in large quantities. It is also expressed in the lymph nodes
 CC for a better immune response. Sequences AA87212-287217 represent
 CC synthetic DNA sequences encoding BONT Hc fragments used in the present
 CC invention. These were optimised for codon usage for expression in yeast.
 XX
 SQ Sequence 1338 BP; 401 A; 342 C; 249 G; 346 T; 0 other;

Query Match 98.4%; Score 1311; DB 21; Length 1338;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 cgtctgtcttaccttccatcgaatacatcaagaacatcatcaatccatccctcgaac 75
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 15 cgtctgtcttaccttccatcgaatacatcaagaacatcatcaatccatccctcgaac 74
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 76 ctgctgctgaatccatccatccatccatccatccatccatccatccatccatccatcc 135
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 75 ctgctgctgaatccatccatccatccatccatccatccatccatccatccatccatcc 134
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 136 gttcttaagttaacttgcgatccgacgacgaagaatccatccatccatccatccatcc 195
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 135 gttcttaagttaacttgcgatccgacgacgaagaatccatccatccatccatccatcc 194
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 196 tcttccaataatcgaatttccgtgaagaatgcatcgtatccatccatccatccatccatcc 255
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 195 tcttccaataatcgaatttccgtgaagaatgcatcgtatccatccatccatccatccatcc 254
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 256 ttctccacctctctctgataccgataccgataatccatccatccatccatccatccatcc 315
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 255 ttctccacctctctctgataccgataccgataatccatccatccatccatccatccatcc 314
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 316 gaatacaccatcacaacatcgatggaagaacatccgtgttggaagtatctctgaactac 375
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 315 gaatacaccatcacaacatcgatggaagaacatccgtgttggaagtatctctgaactac 374
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 376 ggtgaatcatctggaactctgacgacgacgacgacgacgacgacgacgacgacgacgac 435
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 375 ggtgaatcatctggaactctgacgacgacgacgacgacgacgacgacgacgacgacgac 434
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 436 tactctcagatgatacaacatctctgatacatcaatccgtggaactctctgatacaccac 495
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 435 tactctcagatgatacaacatctctgatacatcaatccgtggaactctctgatacaccac 494
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 496 aacaatcgtctgaatatactaccccaaatctcaatcaatcaatcaatcaatcaatcaatca 555
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 495 aacaatcgtctgaatatactaccccaaatctcaatcaatcaatcaatcaatcaatcaatca 554
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 556 atctccaatctggtatacatccacgctctcaataacatcatctcaatctgaacgagttgt 615
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 555 atctccaatctggtatacatccacgctctcaataacatcatctcaatctgaacgagttgt 614
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 616 cgtgacactcaccgctacatctgatacgaataatctcaatctgtctgacgaagaactgac 675
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 615 cgtgacactcaccgctacatctgatacgaataatctcaatctgtctgacgaagaactgac 674
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 676 gaaaagaatcaaaaactctgtacgacgaacccgctcaatctgtgtctctgaaagcttc 735
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 675 gaaaagaatcaaaaactctgtacgacgaacccgctcaatctgtgtctctgaaagcttc 734
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 736 tgggtgactaccctgacgacgacgaacacgctacatgctgatactgatacgaatccgaac 795
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 735 tgggtgactaccctgacgacgacgaacacgctacatgctgatactgatacgaatccgaac 794
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 796 aaatagcttgaacgctcaacaatgtaggatccgctgtatcatgtactgaaagcttcgct 855
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 795 aaatagcttgaacgctcaacaatgtaggatccgctgtatcatgtactgaaagcttcgct 854
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 856 gttctgttatgactcccaacatctccatccatccatccatccatccatccatccatccatcc 915
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 855 gttctgttatgactcccaacatctccatccatccatccatccatccatccatccatccatcc 914
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 916 atcataaagaatcagctctgttatacgaagaatatacgttgcgaacaatgatactgtata 975
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 915 atcataaagaatcagctctgttatacgaagaatatacgttgcgaacaatgatactgtata 974
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 976 tacatcaatgttgaatataagaacaagaataacgctgtgtatccaaatgtcttcagct 1035
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 975 tacatcaatgttgaatataagaacaagaataacgctgtgtatccaaatgtcttcagct 1034
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1036 ggtgtgaaagaatctctgtctgtctgtggaatcccgagacgttgtgatactgtctcaggtta 1095
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1035 ggtgtgaaagaatctctgtctgtctgtggaatcccgagacgttgtgatactgtctcaggtta 1094
 DB ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1096 gttgtatgaatccaagaacgacgaggtatcatcactaacaatgcaaatgaaatgtgcag 1155

Db 1095 gtctgaatgaatccaagaccagcgatcactacaatcgaatgaatctgcag 1154
 QY 1156 gacacaaatgaatgaatcgcgttcacgcgttcacacagttcacaatctgcga 1215
 Db 1155 gacacaaatgaatgaatcgcgttcacgcgttcacacagttcacaatctgcga 1214
 QY 1216 ctggtgctcccaactgcgcgaatcgcgcgaatcgcgcgaatcgcgcga 1275
 Db 1215 ctggtgctcccaactgcgcgaatcgcgcgaatcgcgcgaatcgcgcga 1274
 QY 1276 tcttggaagtccatcccggttgatgacggttggtgaagtcgcgtga 1326
 Db 1275 tcttggaagtccatcccggttgatgacggttggtgaagtcgcgtga 1325

RESULT 5

AAV30576
ID AAV30576 standard; DNA: 1351 BP.

XX AAV30576;

XX 07-DEC-1998 (first entry)

XX Clostridium botulinum toxin A fragment C gene in pHisBotA(syn).

KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;

KM botulinm; ds.

XX Clostridium botulinum serotype A.

OS Synthetic.

XX Key Location/Qualifiers

XX CDS 1..1338

XX /tag- a

XX MO9808540-A1.

XX 05-MAR-1998.

XX 28-AUG-1997; 97MO-US15394.

XX 28-AUG-1996; 96US-0704159.

XX (OPHI-) OPHIDIAN PHARM INC.

XX Thalley BS, Williams JA;

XX WPI; 1998-230234/20.

XX P-P-SDB; AAM68391.

XX Host cell containing recombinant expression vector encoding

XX Clostridium botulinum type B or E toxin - useful to treat humans

XX and other animals at risk of intoxication with clostridial toxin

XX Example 29; Page 279-281; 428bp; English.

XX This is the DNA sequence of the Clostridium botulinum serotype A

XX toxin C fragment gene contained in plasmid pHisBotA(syn). The

XX encoded toxin A polypeptide (see AAM68391) has a histidine-tagged

XX N-terminal extension. The vector was used to express native

XX (i.e. non-fusion) soluble C fragment in Escherichia coli host

XX cells. The invention relates to recombinant proteins derived from

XX C. botulinum toxins. Methods are provided which allow for the

XX isolation of soluble recombinant proteins free of significant

XX endotoxin contamination. Preferred hosts for production of

XX recombinant proteins are E. coli, insect cells and yeast cells.

XX The recombinant toxins are used as immunogens for the production

XX of vaccines and antitoxins that are useful in the treatment of

XX humans and animals at risk of intoxication with clostridial toxin.

XX Sequence 1351 BP; 407 A; 348 C; 247 G; 349 T; 0 other;

Query Match 98.4%; Score 1311; DB 19; Length 1351;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 cgtctgctgtctactcctcactgaatatacaagaactatcaatcactcactgaac 75
 Db 28 cgtctgctgtctactcctcactgaatatacaagaactatcaatcactcactgaac 87
 QY 76 ctgcgcgaatccaatccaatcactgaatcactgaatcactgaatcactgaatc 135
 Db 88 ctgcgcgaatccaatccaatcactgaatcactgaatcactgaatcactgaatc 147
 QY 136 ggttcataaagttaacttcgaatccgaatcgaatcgaatcgaatcgaatcga 195
 Db 148 ggttcataaagttaacttcgaatccgaatcgaatcgaatcgaatcgaatcga 207
 QY 196 tcttcacaatcgaatgtatccctgaagaagtctcgttatatacaatctatgtcga 255
 Db 208 tcttcacaatcgaatgtatccctgaagaagtctcgttatatacaatctatgtcga 267
 QY 256 tcttcacaatcgaatgtatccctgaagaagtctcgttatatacaatctatgtcga 315
 Db 268 tcttcacaatcgaatgtatccctgaagaagtctcgttatatacaatctatgtcga 327
 QY 316 gaatacacatcatalcaactgcatacgaatgaagaatcgttgaagaatcgtgaact 375
 Db 328 gaatacacatcatalcaactgcatacgaatgaagaatcgttgaagaatcgtgaact 387
 QY 376 ggtgaatcactcgtgactcgtcgaatcgaatcgaatcgaatcgaatcgaatcga 435
 Db 388 ggtgaatcactcgtgactcgtcgaatcgaatcgaatcgaatcgaatcgaatcga 447
 QY 436 tactcagaatgaatcaatcctcgtactacatcgaatcgaatcgaatcgaatcga 495
 Db 448 tactcagaatgaatcaatcctcgtactacatcgaatcgaatcgaatcgaatcga 507
 QY 496 aacacatcgtcgaatcactccaatcactcaatcactcaatcactcaatcactca 555
 Db 508 aacacatcgtcgaatcactccaatcactcaatcactcaatcactcaatcactca 567
 QY 556 atctccaatcgtcgaatcactccaatcactcaatcactcaatcactcaatcactca 615
 Db 568 atctccaatcgtcgaatcactccaatcactcaatcactcaatcactcaatcactca 627
 QY 616 cgtgacatcactcgtcgaatcactccaatcactcaatcactcaatcactcaatcactca 675
 Db 628 cgtgacatcactcgtcgaatcactccaatcactcaatcactcaatcactcaatcactca 687
 QY 676 gaaagaagaatcgaatcactccaatcactcaatcactcaatcactcaatcactca 735
 Db 688 gaaagaagaatcgaatcactccaatcactcaatcactcaatcactcaatcactca 747
 QY 736 tgggtgtaactcgtcgaatcactccaatcactcaatcactcaatcactcaatcactca 795
 Db 748 tgggtgtaactcgtcgaatcactccaatcactcaatcactcaatcactcaatcactca 807
 QY 796 aatacgttgaatcgaatcactccaatcactcaatcactcaatcactcaatcactca 855
 Db 808 aatacgttgaatcgaatcactccaatcactcaatcactcaatcactcaatcactca 867
 QY 856 ggttcgttgaatcgaatcactccaatcactcaatcactcaatcactcaatcactca 915
 Db 868 ggttcgttgaatcgaatcactccaatcactcaatcactcaatcactcaatcactca 927
 QY 916 atcactgaatcgaatcactccaatcactcaatcactcaatcactcaatcactca 975
 Db 928 atcactgaatcgaatcactccaatcactcaatcactcaatcactcaatcactca 987
 QY 976 tacatcaatgttgaatgaagaagaatcactcgttgaatgaagaagaatcactcgttga 1035
 Db 988 tacatcaatgttgaatgaagaagaatcactcgttgaatgaagaagaatcactcgttga 1047

QY 1036 ggtgtagaaaagatctgtctgtctgtgaatacccgagcgttgtagtaattctgtcacgta 1095
 |||||
 Db 1048 ggtgtagaaaagatctgtctgtctgtgaatacccgagcgttgtagtaattctgtcacgta 1107
 QY 1096 gttgtagtaaatccaagaagcagggatcactaactaactgcaaatgtaattctgcag 1155
 |||||
 Db 1108 gttgtagtaaatccaagaagcagggatcactaactaactgcaaatgtaattctgcag 1167
 QY 1156 gacaacaatgtagaagatctgtctgtctgtgaatacccgagcgttgtagtaattctgtcacgta 1215
 |||||
 Db 1168 gacaacaatgtagaagatctgtctgtctgtgaatacccgagcgttgtagtaattctgtcacgta 1227
 QY 1216 ctgtgtctctccaactgtagaactcgtcagatcgaacgtctctctcgcacactgtggttcg 1275
 |||||
 Db 1228 ctgtgtctctccaactgtagaactcgtcagatcgaacgtctctctcgcacactgtggttcg 1287
 QY 1276 tcttggaggttcatcccggttgtagtaacggttgggttgaacgttcgcgtgtaa 1326
 |||||
 Db 1288 tcttggaggttcatcccggttgtagtaacggttgggttgaacgttcgcgtgtaa 1338

RESULT 6
 AAT29246
 ID AAT29246 standard; DNA; 1402 BP.
 XX
 AC AAT29246;
 XX
 DT 07-JUL-1996 (first entry)
 XX
 DE Type A neurotoxin C fragment-polyhistidine tag gene fusion.
 XX
 KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
 KW Clostridium botulinum; polyhistidine; vector; PETHisa; pHISBot; ds.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1317
 FT /*tag= a
 FT /product= pHISBot fusion protein
 XX
 PN W09612802-A1.
 XX
 PD 02-MAY-1996.
 XX
 PF 23-OCT-1995; 95WO-US13737.
 XX
 PR 07-JUN-1995; 95US-0480604.
 PR 24-OCT-1994; 94US-0329154.
 PR 16-MAR-1995; 95US-0405496.
 PR 14-APR-1995; 95US-0422711.
 XX
 PA (OPHI-) OPHIDIAN PHARM INC.
 XX
 PI Pirca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
 PI Williams JA;
 XX
 DR WPI; 1996-230603/23.
 DR P-PSDB; AAR95009.
 XX
 PT Fusion proteins comprising non-toxin protein and part of toxin
 PT useful to form anti-toxins against Clostridium botulinum type A, and
 PT C. difficile type toxins, and to treat C. difficile intoxication,
 PT partic. diarrhoea
 XX
 PS Example 24; Page 340-342; 434pp; English.
 CC A nucleotide sequence (AAT29246) present in vector pETHisa encodes
 CC the pHISBot fusion protein (AAR95009) comprising a polyhistidine
 CC affinity tag and fragment C (see also AAR95008) of the Clostridium
 CC botulinum type A neurotoxin. The pHISBot protein was expressed
 CC in Escherichia coli as a soluble protein and was purified by
 CC metal chelate affinity chromatography to obtain a product free

CC of endotoxin contamination that may be useful as an immunogen
 CC in vaccine comps.
 XX
 SQ Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other;

Query Match 98.4%; Score 1311; DB 17; Length 1402;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 cgtctgtcttctacccttccatgaatacatcaaatcatcaatccatccatccatgaac 75
 |||||
 Db 79 cgtctgtcttctacccttccatgaatacatcaaatcatcaatccatccatccatgaac 138
 QY 76 ctggtgtagaataccaataaccctgatacgcgtgtctcgttaagcttccaaataatcaatc 135
 |||||
 Db 139 ctggtgtagaataccaataaccctgatacgcgtgtctcgttaagcttccaaataatcaatc 198
 QY 136 gttctcaaggttaactcgcgatccgacgaagaatcagatccagctgttcaactgga 195
 |||||
 Db 199 gttctcaaggttaactcgcgatccgacgaagaatcagatccagctgttcaactgga 258
 QY 196 tcttccaaaatcgaagttaactcctgaagaatgctatcgtatataactctatgttaaga 255
 |||||
 Db 259 tcttccaaaatcgaagttaactcctgaagaatgctatcgtatataactctatgttaaga 318
 QY 256 tcttccaaactctcttgatcgatccggaataacttcaactccatctctctgaacat 315
 |||||
 Db 319 tcttccaaactctcttgatcgatccggaataacttcaactccatctctctgaacat 378
 QY 316 gaatacacatcatcaacgcatggaagaacaaattcgttggaaagtatctcgtgaactac 375
 |||||
 Db 379 gaatacacatcatcaacgcatggaagaacaaattcgttggaaagtatctcgtgaactac 438
 QY 376 ggtgaaatcatctggaactctggaagcactcagaatcaacaacgctgttattcaaa 435
 |||||
 Db 439 ggtgaaatcatctggaactctggaagcactcagaatcaacaacgctgttattcaaa 498
 QY 436 tactctcagaatgtaacactctctgatacatcaactcgttgggttctgttaccatcac 495
 |||||
 Db 499 tactctcagaatgtaacactctctgatacatcaactcgttgggttctgttaccatcac 558
 QY 496 aacaatcgtctgaataactccaanaatcatcaatcaacgctgttctgaaccagaacg 555
 |||||
 Db 559 aacaatcgtctgaataactccaanaatcatcaatcaacgctgttctgaaccagaacg 618
 QY 556 atctccaatctgggtgaataccaacgcttctaataacatcatgttcaactgtagggt 615
 |||||
 Db 619 atctccaatctgggtgaataccaacgcttctaataacatcatgttcaactgtagggt 678
 QY 616 cgtgacactcaccgctacatctggaataactcaatctgttggacaagaagctggaac 675
 |||||
 Db 679 cgtgacactcaccgctacatctggaataactcaatctgttggacaagaagctggaac 738
 QY 676 gaaaagaatcaaaagactgtttagcagaacacagctccaattctgttctcgtgaagactc 735
 |||||
 Db 739 gaaaagaatcaaaagactgtttagcagaacacagctccaattctgttctcgtgaagactc 798
 QY 736 tgggtgtagtaactcgtcagtagaagaacacgcttactatcatgttgaactgtgagatccgaac 795
 |||||
 Db 799 tgggtgtagtaactcgtcagtagaagaacacgcttactatcatgttgaactgtgagatccgaac 858
 QY 796 aatagctgtagcgtcaaatgttagtctcgcgttgaatgaatgaatgaatgaatgaatgaat 855
 |||||
 Db 859 aatagctgtagcgtcaaatgttagtctcgcgttgaatgaatgaatgaatgaatgaatgaat 918
 QY 856 ggttctgttagtaactcaacatctacctgaactctcctgttgcgtgtgtacaaatcc 915
 |||||
 Db 919 ggttctgttagtaactcaacatctacctgaactctcctgttgcgtgtgtgtacaaatcc 978
 QY 916 atcatcaagaataacgctctgttgaagaagaacatatactgttcgaacaatgtagtga 975
 |||||
 Db 979 atcatcaagaataacgctctgttgaagaagaacatatactgttcgaacaatgtagtga 1038

XX	New nucleic acids encoding the carboxy- or amino-terminal portions of		
PT	the heavy chain of botulinum neurotoxin of serotype A-G, useful as		
PT	vaccine against botulism		
XX			
PS	Disclosure; Fig 2a; 73pp; English.		
XX			
CC	Botulinum neurotoxins are translated as a single 150 kDa polypeptide		
CC	chain and then posttranslationally nicked, forming a dichain		
CC	consisting of a 100 kDa heavy chain and a 50 kDa light chain which		
CC	remain linked by a disulfide bond. Nucleic acids encoding the		
CC	carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy		
CC	chain of botulinum neurotoxin (BoNT) can be used in recombinant		
CC	expression vectors and expressed in transformed cells to produce		
CC	peptide antigens useful for eliciting an immune response to give		
CC	protective immunity against botulinum neurotoxin, which causes		
CC	botulism. The nucleic acids are expressible in a recombinant		
CC	organisms such as Escherichia coli or Pichia pastoris. The use		
CC	of recombinant nucleic acids are advantageous since it eliminates		
CC	the need to culture large quantities of hazardous toxin-producing		
CC	bacteria. Production yield from the genetically engineered product		
CC	is also high and cost of production is lower. The nucleic acids can		
CC	be derived from Clostridium botulinum serotypes A-G.		
XX			
XX			
SO	Sequence 1323 BP; 404 A; 334 C; 242 G; 343 T; 0 other;		
Query Match	98.4%; Score 1310.4; DB 22; Length 1323;		
Best Local Similarity	99.9%; Pred. No. 0;		

QY	21	gctgtctaccttcaatgtaatacatcaagaacatcatcaactccatccatctgaaactg	80
Db	12	gattgtctaccttcaatgtaatacatcaagaacatcatcaactccatccatctgaaactg	71
QY	81	ctacgaatccaatcaactcgatcgcgctctctgcgacgcttccaaatcaaatcggttc	140
Db	72	ctacgaatccaatcaactcgatcgcgctctctgcgacgcttccaaatcaaatcggttc	131
QY	141	taaagttaacttcgatccgatcgcgacagaatcatgatccagctgtgttcaatctgaaatcttc	200
Db	132	taaagttaacttcgatccgatcgcgacagaatcatgatccagctgtgttcaatctgaaatcttc	191
QY	201	caaaatcgaaattatcccggaagaatgycatctgtatcaactctatgtgaagaaacttcc	260
Db	192	caaaatcgaaattatcccggaagaatgycatctgtatcaactctatgtgaagaaacttcc	251
QY	261	caactctctctgatatcgatcccgaaatattcaactlccatctctctgaaacatgata	320
Db	252	caactctctctgatatcgatcccgaaatattcaactlccatctctctgaaacatgata	311
QY	321	cacccttcacatgcatactggaagaacaattctggtttggaagatctctgaaactaggtta	380
Db	312	cacccttcacatgcatactggaagaacaattctggtttggaagatctctgaaactaggtta	371
QY	381	aatcatctcgaactctcgacagacactccaggaatactcaaacagcgltgtgtatactcaatatttc	440
Db	372	aatcatctcgaactctcgacagacactccaggaatactcaaacagcgltgtgtatactcaatatttc	431
QY	441	tcaagtatgtaacaatctcttgatacatcaatctgcgtcggatcttcggttaccatcaaccaaa	500
Db	432	tcaagtatgtaacaatctcttgatacatcaatctgcgtcggatcttcggttaccatcaaccaaa	491
QY	501	tgcgtcgtataaactccaatctacatcaaacgcgctctgatacgcgacgaagaaacgatacttc	560
Db	492	tgcgtcgtataaactccaatctacatcaaacgcgctctgatacgcgacgaagaaacgatacttc	551
QY	561	caatctgggtatacatccaacgcttcttaataacaatcatttcaactcggacggtttgctgtta	620
Db	552	caatctgggtatacatccaacgcttcttaataacaatcatttcaactcggacggtttgctgtta	611
QY	621	caatcaccgctacatctgataccaatattctcaatctgttcgacaagaactgaaacgaaa	680

```

Db 612 cactaccgctacacttgatcaataactcaatctgttcgcagaagaactgaacgaaa 671
Qy 681 agaatcaagaactctgtacacacacgctccaaattctgtatccctgaagaactctggg 740
Db 672 agaatcaagaactctgtacacacacgctccaaattctgtatccctgaagaactctggg 731
Qy 741 tgaactctcgtacgaacacacgctccaaattctgtatccctgaagaactctggg 800
Db 732 tgaactctcgtacgaacacacgctccaaattctgtatccctgaagaactctggg 791
Qy 801 cgttgaactcaagaactctgtacacacgctccaaattctgtatccctgaagaactctggg 860
Db 792 cgttgaactcaagaactctgtacacacgctccaaattctgtatccctgaagaactctggg 851
Qy 861 tgaactctcgtacgaacacacgctccaaattctgtatccctgaagaactctggg 920
Db 852 tgaactctcgtacgaacacacgctccaaattctgtatccctgaagaactctggg 911
Qy 921 caagaataacgctctgtacgaacacacgctccaaattctgtatccctgaagaactctggg 980
Db 912 caagaataacgctctgtacgaacacacgctccaaattctgtatccctgaagaactctggg 971
Qy 981 caatgtgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1040
Db 972 caatgtgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1031
Qy 1041 agaaaagaactctgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1100
Db 1032 agaaaagaactctgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1091
Qy 1101 aatgaataacgctctgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1160
Db 1092 aatgaataacgctctgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1151
Qy 1161 caatgtgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1220
Db 1152 caatgtgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1211
Qy 1221 tgaactctcgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1280
Db 1212 tgaactctcgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1271
Qy 1281 ggaactctcgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1332
Db 1272 ggaactctcgtacgaacacacgctccaaattctgtatccctgaagaactctggg 1323

```

RESULT 9
 ID AAS4484
 AAS4484 standard; DNA; 1326 BP.
 AC AAS4484;
 DT 11-APR-2001 (first entry)
 XX
 DE Botulinum toxin heavy chain C-terminal coding sequence (serotype A).
 KW Botulinum; toxin; neurotoxin; heavy chain; recombinant expression;
 KW recombinant vector; antigen; immune response; vaccine; bacterium;
 KW infection; ds.
 XX
 OS Synthetic.
 OS Clostridium botulinum.
 XX
 FH Key Location/Qualifiers
 FT CDS 13..1320
 FT /tag= a
 FT /product= H_C peptide fragment
 XX
 XX WO200067700-A2.
 XX 16-NOV-2000.

```

PF 12-MAY-2000; 2000MO-US12890.
XX
XX 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
PI
XX
XX MPI: 2001-016048/02.
DR P-PSDB: AAB04090.
XX
XX New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulinism
XX
XX Disclosure; Fig 3a; 73pp; English.
XX
XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a di-chain which
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulinism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.
XX
XX Sequence 1326 BP; 404 A; 336 C; 243 G; 343 T; 0 other;
SQ

```

Query Match 98.2%; Score 1308.2; DB 22; Length 1326;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1310; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy 20 tgcgtctacactgaatcaataacatcaataactcactcgaactgc 79
Db 14 tgcgtctacactgaatcaataacatcaataactcactcgaactgc 73
Qy 80 gctacgaatcaatcactgacactgtctcgtacgctccaaatacaatcgtgt 139
Db 74 gctacgaatcaatcactgacactgtctcgtacgctccaaatacaatcgtgt 133
Qy 140 ctaagttaactcgtacgacatcgacagaagaatcgatccgagctgttaactcgtgaactc 199
Db 134 ctaagttaactcgtacgacatcgacagaagaatcgatccgagctgttaactcgtgaactc 193
Qy 200 ccaaatcgaagtatccttaagaatgctatcgtatacaactcgtatgaagaactcct 259
Db 194 ccaaatcgaagtatccttaagaatgctatcgtatacaactcgtatgaagaactcct 253
Qy 260 ccaactccttcgtacgctatcccgaaataactcaactcactcctcgtgaagaat 319
Db 254 ccaactccttcgtacgctatcccgaaataactcaactcactcctcgtgaagaat 313
Qy 320 acacatcatcaactgacatggaagaacatctcgttggaaagtatcctcgaactcgtgt 379
Db 314 acacatcatcaactgacatggaagaacatctcgttggaaagtatcctcgaactcgtgt 373
Qy 380 aatcatcgtactctgcaagacactcgaagaatacaagaagcgtgtgttcaactact 439

```

Db 374 aaatcatctgagctctgaggaacatacagaacagcggtgtgtatctcaact 433
 QY 440 ctacagatcatcaatctctgactacatcatcgctgactcttgatcatcaccaca 499
 Db 434 ctacagatcatcaatctctgactacatcatcgctgactcttgatcatcaccaca 493
 QY 500 atcgctgaatactccaatactacatcaacgcgctgctgactcgacagaacagact 559
 Db 494 atcgctgaatactccaatactacatcaacgcgctgactcgacagaacagactct 553
 QY 560 ccaatcgggttaacatccacgcttctaatacatcagtccaactggaagctgtgtg 619
 Db 554 ccaatcgggttaacatccacgcttctaatacatcagtccaactggaagctgtgtg 613
 QY 620 acactcaacgctcatctgtgatacactcaactcaactctgttcgcaaaaagaactggaaga 679
 Db 614 acactcaacgctcatctgtgatacactcaactcaactctgttcgcaaaaagaactggaaga 673
 QY 680 aagaatcaaaagactgtgacacacacagctccaatttcgttalcctgaagaactctg 739
 Db 674 aagaatcaaaagactgtgacacacacagctccaatttcgttalcctgaagaactctg 733
 QY 740 gtaactactcgcagtgacacacacagctcatcactgctgaactgtgtacgacagaacat 799
 Db 734 gtaactactcgcagtgacacacacagctcatcactgctgaactgtgtacgacagaacat 793
 QY 800 acgttgacgtcaacaagtgtgatacgcggttatacgttaccgtgaagaagctcggtgtgt 859
 Db 794 acgttgacgtcaacaagtgtgatacgcggttatacgttaccgtgaagaagctcggtgtgt 853
 QY 860 ctgttactgacacacacatctactgactcttccctgtacacgctgacacacacatctca 919
 Db 854 ctgttactgacacacacatctactgactcttccctgtacacgctgacacacacatctca 913
 QY 920 tcaagaatacgcgtctgcgtgaacagacacacatctgttcgaacatgactcggtatataca 979
 Db 914 tcaagaatacgcgtctgcgtgaacagacacacatctgttcgaacatgactcggtatataca 973
 QY 980 tcaatgtgtgagttaagaacaaagaatacactgctgtgtactacaatgcttccagctgtgtg 1039
 Db 974 tcaatgtgtgagttaagaacaaagaatacactgctgtgtactacaatgcttccagctgtgtg 1033
 QY 1040 tagaanaagactgtctgtctctcgtgaacatcccgagacgtgtgttaactgttccaggtatgtg 1099
 Db 1034 tagaanaagactgtctgtctctcgtgaacatcccgagacgtgtgttaactgttccaggtatgtg 1093
 QY 1100 taatgaataccaaagaacagacagcggtgtactactaaacatgcaaatgaaatctgcagagaca 1159
 Db 1094 taatgaataccaaagaacagacagcggtgtactactaaacatgcaaatgaaatctgcagagaca 1153
 QY 1160 acaatgtaacgatatcggttccatcggttccacagcttccacaaatatacgttaactgtg 1219
 Db 1154 acaatgtaacgatatcggttccatcggttccacagcttccacaaatatacgttaactgtg 1213
 QY 1220 ttgcttcaacactggtataactcgtacagtcgaacgcttccctcgcgactcgtgtgtctct 1279
 Db 1214 ttgcttcaacactggtataactcgtacagtcgaacgcttccctcgcgactcgtgtgtctct 1273
 QY 1280 gggagttcatcccggtgtgtagagcggtgtggtgaacgtccgctgttaagaattc 1332
 Db 1274 gggagttcatcccggtgtgtagagcggtgtggtgaacgtccgctgttaagaattc 1326
 RESULT 10
 AAA54588
 ID AAA54588 standard; DNA; 1332 BP.
 AC AAA54588;
 XX
 XX 11-APR-2001 (first entry)
 DT
 DE Sequence encoding botulism toxin C fragment (serotype A).
 XX

KM Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
 KM recombinant vector; antigen; immune response; vaccine; bacterium;
 KM Infection; ds.
 OS Synthetic.
 OS Clostridium botulinum.
 PN WO20067700-A2.
 XX
 PD 16-NOV-2000.
 XX
 PF 12-MAY-2000; 2000WO-US12890.
 XX
 PR 12-MAY-1999; 99US-0133865.
 PR 12-MAY-1999; 99US-0133866.
 PR 12-MAY-1999; 99US-0133867.
 PR 12-MAY-1999; 99US-0133868.
 PR 12-MAY-1999; 99US-0133869.
 PR 12-MAY-1999; 99US-0133870.
 PR 29-JUL-1999; 99US-0146192.
 XX
 PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
 XX
 PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
 PI
 XX
 DR WPI; 2001-016048/02.
 XX
 PT New nucleic acids encoding the carboxy- or amino-terminal portions of
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
 PT vaccine against botulism
 PS
 PS
 PS Example 7; Page 37; 73pp; English.
 XX
 CC Botulism neurotoxins are translated as a single 150 kDa polypeptide
 CC chain and then posttranslationally nicked, forming a di-chain
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
 CC remain linked by a disulfide bond. Nucleic acids encoding the
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
 CC expression vectors and expressed in transformed cells to produce
 CC peptide antigens useful for eliciting an immune response to give
 CC protective immunity against botulinum neurotoxin, which causes
 CC botulism. The nucleic acids are expressible in a recombinant
 CC organisms such as Escherichia coli or Pichia pastoris. The use
 CC of recombinant nucleic acids are advantageous since it eliminates
 CC the need to culture large quantities of hazardous toxin-producing
 CC bacterium. Production yield from the genetically engineered product
 CC is also high and cost of production is lower. The nucleic acids can
 CC be derived from Clostridium botulinum serotypes A-G.
 XX
 SQ Sequence 1332 BP; 401 A; 342 C; 244 G; 344 T; 1 other;
 Query Match 94.3%; Score 1256.6; DB 22; Length 1332;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 1303; Conservative 1; Mismatches 1; Indels 6; Gaps 4;
 QY 16 cgtctgtgtctactcactgaatacagaacatcaatactcatctcaac 75
 Db 15 cgtctgtgtctactcactgaatacagaacatcaatactcatctcaac 74
 QY 76 ctgcgtacgatacacaatcactatcgactcgtcgtcgtacgttccaataatc 135
 Db 75 ctgcgtacgatacacaatcactatcgactcgtcgtcgtacgttccaataatc 134
 QY 136 ggtctaaagttaactcgcgacgcgacgacaagaatcagatccagctgttcaatcggaa 195
 Db 135 ggtctaaagttaactcgcgacgcgacgacaagaatcagatccagctgttcaatcggaa 194
 QY 196 tcttcaaaaatcgaagtatcttgaagaatcgtatcgtatacaactatgttaagaac 255
 Db 195 tcttcaaaaatcgaagtatcttgaagaatcgtatcgtatacaactatgttaagaac 254

QY	256	ttccacccctcttcggaatccgataccggaataacttaactcaatcattctctgaacat	315
Db	255	ttccacccctcttcggaatccgataccggaataacttaactcaatcattctctgaacat	313
QY	316	gaatacacatcatcaactctgatactggaanaaactctcgtctggaagatctctgaactac	375
Db	314	gaatacacatcatcaactctgatactggaanaaactctcgtctggaagatctctgaactac	373
QY	376	ggtgaaatactctcggactctcgagagacacttcaggaaataacaacagcgtgtgattccaa	435
Db	374	ggtgaaatactctcggactctcgagagacacttcaggaaataacaacagcgtgtgattccaa	433
QY	436	tactctcagaatgatacaacatctctgatacatcaatctgcgtgatatctctgataccacc	495
Db	434	tactctcagaatgatacaacatctctgatacatcaatctgcgtgatatctctgataccacc	493
QY	496	aacaaatctgtctgataatacttcaaaactctatactcaagcgcgtctgatacgacagaacacg	555
Db	494	aacaaatctgtctgataatacttcaaaactctatactcaagcgcgtctgatacgacagaacacg	552
QY	556	atccccaactctgggttaacatccacacgcgtcttataatacatcttcaaaactgagacgttgt	615
Db	553	atccccaactctgggttaacatccacacgcgtcttataatacatcttcaaaactgagacgttgt	611
QY	616	cgtgaaactctacgcgtatactctgatactcaaaataactcaatctgtctcgacaagaactgaac	675
Db	612	cgtgaaactctacgcgtatactctgatactcaaaataactcaatctgtctcgacaagaactgaac	671
QY	676	gaaaaagaatactcaaaagccgtgtacgcgaacacgcgtccaactctcgtgtatccctgaagaacttc	735
Db	672	gaaaaagaatactcaaaagccgtgtacgcgaacacgcgtccaactctcgtgtatccctgaagaacttc	731
QY	736	tgggtgactactacgtcgatcgcgaacaaacgctactacatgctgaaactctgtaacgtccgac	795
Db	732	tgggtgactactacgtcgatcgcgaacaaacgctactacatgctgaaactctgtaacgtccgac	791
QY	796	aaatacagcttaacgttcaacaaatgtaggtatctccgcgtgtacatgttaactctgaagagttccggt	855
Db	792	aaatacagcttaacgttcaacaaatgtaggtatctccgcgtgtacatgttaactctgaagagttccggt	851
QY	856	ggtctctgtatgatactcaacacatctacacgttgaactctctccctgtacacgtgtgataccaatctc	915
Db	852	ggtctctgtatgatactcaacacatctacacgttgaactctctccctgtacacgtgtgataccaatctc	911
QY	916	atcatacagaatacagcgtctctgtttaacaagaagacataatcgtctcgacaacatgatactgtta	975
Db	912	atcatacagaatacagcgtctctgtttaacaagaagacataatcgtctcgacaacatgatactgtta	968
QY	976	tacatacaatctttagttaagaacaaagaataacgcgtctcgtctacaaatgattcttaagct	1035
Db	969	tacatacaatctttagttaagaacaaagaataacgcgtctcgtctacaaatgattcttaagct	1028
QY	1036	ggtgttaagaanaagatctctgtctctcgtcgtggaataatcccggaacgttgtgtaactctctcaagta	1095
Db	1029	ggtgttaagaanaagatctctgtctctcgtcgtggaataatcccggaacgttgtgtaactctctcaagta	1088
QY	1096	gctgttaatgaataatcccaagaacgcagcgggtatacctaactaacaatgcaaaatgaattctgcag	1155
Db	1089	gctgttaatgaataatcccaagaacgcagcgggtatacctaactaacaatgcaaaatgaattctgcag	1148
QY	1156	gacacaaatggtatacgaatacgcgttttccatcgcgttttccaccagtttaacaatatcgtctaa	1215
Db	1149	gacacaaatggtatacgaatacgcgttttccatcgcgttttccaccagtttaacaatatcgtctaa	1208
QY	1216	ctggttgccttccaactctgttaacaatctgttcagatcgaacgcttctctcgcgaactctggtgtgc	1275
Db	1209	ctggttgccttccaactctgttaacaatctgttcagatcgaacgcttctctcgcgaactctggtgtgc	1268
QY	1276	tcttggaagattcaatcccggtgtgtatgcaggtttgggtgtgaacgtccgcgtgtaa	1326
Db	1269	tcttggaagattcaatcccggtgtgtatgcaggtttgggtgtgaacgtccgcgtgtaa	1319

ID	AAV30575	standard; DNA; 1546 BP.
XX	AAV30575;	
XX	07-DEC-1998	(first entry)
XX	Clostridium botulinum toxin A fragment C gene in pHisBoTA.	
XX	Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;	
XX	botulism; ds.	
XX	Clostridium botulinum serotype A.	
XX	Synthetic.	
XX	Key	Location/Qualifiers
XX	FT CDS	108..1496
XX	FT	/*tag= a
XX	W09808540-A1.	
XX	05-MAR-1998.	
XX	28-AUG-1997;	97MO-US15394.
XX	28-AUG-1996;	96US-0704159.
XX	(OPHI-) OPHIDIAN PHARM INC.	
XX	Thalley BS, Williams JA;	
XX	WPI: 1998-230234/20.	
XX	P-PSDB: AAM68390.	
XX	Host cell containing recombinant expression vector encoding Clostridium botulinum type B or E toxin - useful to treat humans and other animals at risk of intoxication with clostridial toxin	
XX	Example 28; Page 277-278; 428bp; English.	
XX	This is the DNA sequence of the Clostridium botulinum serotype A toxin C fragment gene contained in plasmid pHisBoTA. The encoded toxin A polypeptide (see AAM68390) has a histidine-tagged N-terminal extension. The vector was used to express native (i.e. non-fusion) soluble C fragment in Escherichia coli host cells. The invention relates to recombinant proteins derived from C. botulinum toxins. Methods are provided which allow for the isolation of soluble recombinant proteins free of significant endotoxin contamination. Preferred hosts for production of recombinant proteins are E. coli, insect cells and yeast cells. The recombinant toxins are used as immunogens for the production of vaccines and antitoxins that are useful in the treatment of humans and animals at risk of intoxication with clostridial toxin.	
XX	Sequence 1546 BP; 629 A; 163 C; 256 G; 498 T; 0 other;	
XX	Query Match	54.6%; Score 727.6; DB 19; Length 1546;
XX	Best Local Similarity	72.2%; Pred. No. 1.4e-194;
XX	Matches 946; Conservative 0; Mismatches 364; Indels 0; Gaps 0;	
XX	23	tgctacaccttcgaataacatcaagaacatcaataacatccatccgaacctgcgt 82
XX	193	tatcaacatttcgtaatatataagaataatataataactctatattgtaattgaat 252
XX	83	acgaatcaatcaacctgatacgactcgtcgtctacgctccaaacacaacatcggtcta 142
XX	253	atgaagaatcaatcaattataagcttactcaggtatcgcatacaaaaataataattgtagta 312
XX	143	aagtaacttcgattccgatacgacaagaatccgattccagctggttcaattcgaaattcca 202
XX	313	aagtaaatcttgcataatagataaaaataacattcaattatattaaattgaagaagttagta 372

QY 203 aatcgaagtataccggaagaatgctacgtatatacaactctatgtaacgaactctcca 262
 Db 373 aatggaagttaattttaaaaaatgctatgtatataataatagatgtaataatlttagta 432
 QY 263 cctcctctgtagtcggtatcccggaataactccaactcctcctctgtaacaatgataca 322
 Db 433 ctagtcttctgtagtaagaatcccttaagatatttaacagatataagcttaataatgata 492
 QY 323 ccatacaactatgctatgtaaaaaaactctgtgtggaagaatctctgtaactgaggtgaa 382
 Db 493 caataataaattgtatggaataaataatcagatggaagaatcacttaattatggtgaa 552
 QY 383 tcatctgactctgaggaactcaggaatacaacagcggtgtgtattcaactatctc 442
 Db 553 taactctgacttatacagataactcaggaataaacaagaatgatttttaataacagtc 612
 QY 443 agatgatcaacatcctgtaactacatcaatcgtgtgactctgttaccatacacaacatc 502
 Db 613 aatgtattatataatcagattatataaacaagatgattttgtaactataataata 672
 QY 503 gtcgtgaataactccaataactacatcaacggtcgtctgtagacacagaacagatctca 562
 Db 673 gattaaataaactcaaaatcatacaatlggaagaatagatcaaaaacacatttcaaa 732
 QY 563 atctggttaacatccacgctctcaataacatcatgttcaaacgtgaggtgtcgtgaca 622
 Db 723 atttaggaatataatcatctgtagtaataataatgatttaataatgattgtgtagata 792
 QY 623 ctacacgtacatctgtagtaataatcactcaatctgtctgcacaagaactgtaaaaaag 682
 Db 793 cactagatataatctgtaataaataatttaactcttctgtagaagaataatgaaaaag 852
 QY 683 aatcaaaagacctgtagcaaacacgctcaactcgtgtagtactgtagaagaactcttggtg 742
 Db 853 aatcaaaagatttataatgataatcaatcaacatccaggtatctttaaagaactcttggtg 912
 QY 743 actacgtcagtagcagcaaacacgctacatcgtctgtagacgtcagcaacaatccg 802
 Db 913 attattacataatgataaaccatactatgtttaaatttataatgataccaataatg 972
 QY 803 ttgacgtcaacaatgtagtactcgtcggtttacatgtaaccgtgaaggtccgtgtctg 862
 Db 973 tctgattgataaattgtagttagttaggttattatctttaaaggcctagaagtagcg 1032
 QY 863 ttatgactaccaacatcactgtaactctcctctgtagcgtgtagtgaacaaattcatca 922
 Db 1033 taatgactaccaacatcattatcaatccaagttgtatagggggaacaaatttattataa 1092
 QY 923 agaaatacgtctgtagaacaagaacatcgtctgcgaacaatgctgtagatataca 982
 Db 1093 aaaaatagctctgtagaataaagaataatgttagaataatgtagtgcgtgtagatata 1152
 QY 983 atgtgtgtagtgaagaacaagaataacgtctggtactacaatgctctcaaggtgtgtag 1042
 Db 1153 atgtgtgtagtgaagaataaagaataataggttagctcaactacacagcagcgtag 1212
 QY 1043 aaaaatctgctgctgtagaatacccgagcgtgtgtagtaatcgtctcaggtgtagttaa 1102
 Db 1213 aaaaatctagtagcatagtaaaataaactgtagttaggaataatcaagtaagtagtaa 1272
 QY 1103 tgaatacgaagaacgacaggtgtagtactaacaatgcaaaatgtagtgcagggcaaca 1162
 Db 1273 tgaagtcaaaaaatgataagaagaatacaaaataaagtaaaatgatttacaagtagta 1332
 QY 1163 atgtgtaagatagtagttagttagttagttagttagttagttagttagttagttagt 1222
 Db 1333 atgtgtaagatagtagttagttagttagttagttagttagttagttagttagttagt 1392
 QY 1223 ctccaactgtagttagttagttagttagttagttagttagttagttagttagttagt 1282
 Db 1393 caagttaattgtagttagttagttagttagttagttagttagttagttagttagttagt 1452

QY 1283 agttacccggttagtagcaggttggtgtagaacgtccgctgtagaagaatc 1332
 Db 1453 aattatctcgttagtagtagttagttagttagttagttagttagttagttagttagt 1502

RESULT 12
 AAC64582
 ID AAC64582 standard; cDNA: 4835 BP.
 XX
 AAC64582;
 DT 15-FEB-2001 (first entry)
 DE BOTM/A neurotoxin prototoxin nucleotide sequence SEQ ID NO:10.
 KW Human; procholecystokinin; CCK A receptor; CCK B receptor;
 KW pancreatitis; antiinflammatory; ss.
 OS Clostridium botulinum.
 PN WO20061192-A2.
 PD 19-OCT-2000.
 PF 06-APR-2000; 2000WO-US09142.
 PR 08-APR-1999; 99US-0288326.
 PA (ALLR) ALLEGAN SALES INC.
 PI Steward LE, Sachs G, Aoki KR;
 DR WPI: 2000-679416/66.
 PT New composition for treating acute pancreatitis, comprises a pancreatic
 cell surface marker binding element, a translocation element that
 transfers polypeptide across vesicular membrane, and a therapeutic
 element -
 PS Disclosure: Page 29-32; 50pp; English.
 CC The present invention describes a composition (I) for treating acute
 CC pancreatitis. (I) comprises a first element containing a binding element
 CC that binds to a pancreatic cell surface marker, a second element
 CC containing a translocation element that facilitates polypeptide transfer
 CC across the vesicular membrane, and a third element containing a
 CC therapeutic element that inhibits enzyme secretion in pancreatic cell
 CC cytoplasm. Also described is a method for making a therapeutic
 CC polypeptide having a binding element selective for cholecystokinin (CCK)
 CC receptor by expressing within a host cell a recombinant chimeric
 CC polypeptide comprising an extein containing a therapeutic element and a
 CC translocational element, and an intein located to the carboxy terminal
 CC of extein having a cysteine, serine or threonine at its amino terminus,
 CC and contacting the extein with a synthetic peptide comprising a CCK
 CC amino acid sequence containing an amidated phenylalanine at a natural
 CC C-terminus, and a cysteine, serine or threonine at its N-terminus, and
 CC a nucleophilic reagent able to cause cleavage of the intein to form a
 CC peptide bond between the extein C-terminus and synthetic peptide
 CC N-terminus through the formation of an activated ester or thio ester
 CC intermediate. (II) has antiinflammatory activity and prevents accumulation
 CC of pancreatic digestive enzymes, and prevents exocytic fusion of vesicles
 CC containing secretory enzymes of pancreas. (I) is useful for treating
 CC acute pancreatitis. The present sequence encodes the Clostridium
 CC botulinum BOTM/A neurotoxin prototoxin which is given in the
 CC exemplification of the present invention.
 SQ Sequence 4835 BP: 1934 A; 517 C; 756 G; 1628 T; 0 other;

Query Match 54.6%; Score 727.6; DB 21; Length 4835;
 Best Local Similarity 72.2%; Pred. No. 2.4e-194;
 Matches 946; Conservative 0; Mismatches 364; Indels 0; Gaps 0;

QY 23 tgcctcactgaatgatacagaacatcaatcactccactcctgaactcgcgt 82
 DB 2945 taccacattactgataatgataatatttaactctcactataatgaat 3004
 QY 83 agcaatccactccactgactcgtctcgtcagcgttccaaataatcaactcgtccta 142
 DB 3005 atgaaagtaactcacttaactgactctcagtagatgatacaaaaataatcgtgtagta 3064
 QY 143 aagttaactcgcgtccgatacgacaagaatagatccagcgttccaaactcgtccta 202
 DB 3065 aagtaaatcttgatcccaatagataaataaataatcactatttaatttagaagaagttagta 3124
 QY 203 aactcgaagtatccctcgaagaatgtagatgatacaaacctcgtatgtagcaaaactcctca 262
 DB 3125 aacttgaggaatattttaaataatgctatgtagtaaaatagtagatgtagaataatttagta 3184
 QY 263 cctcctcctgagatccgataccgaataactcactccactcctcctcgaataatgataca 322
 DB 3185 ctgagctcttgagataagaatcccaagataatttcaacagtagaagcttaaaataatgaaatata 3244
 QY 323 ccaatcaactcagatgaaataaacaatctgcttggaagaatcctcgtcaactcgtgtaaa 382
 DB 3245 caataataaattgtagaataaataatcagatgaggaagaatcacttaattatgtagtaaa 3304
 QY 383 tcaactgagactcgtcagagacactcgaagaatacaaacagcgtgtagtattcaaatcctc 442
 DB 3305 taactcgtactctacagagatcaccgaataaacaagaagtagtattttaaatacagctc 3364
 QY 443 agtgtgacacatctcgtgatacatgaatcgtcgtgactcctcgttaccataccaacatc 502
 DB 3365 aaatgataatatacatgatacatcaatacaacagatgattcttgtaacatacactaataata 3424
 QY 503 gtcgataatcaactcaaaatctacatcaacagcgcgtctgtagcagcagaacagatcctca 562
 DB 3425 gattaaataactcctaaatattatataatgagaagattaaatagatcaaaaacaattctca 3484
 QY 563 atctgagtaacatccacgcgttctaataacatcagatgtaacaaactgagcagctgtagtaca 622
 DB 3485 atttagtaataatcactcagctagataataataatgatttaattagatgtagtagagata 3544
 QY 623 ctacgcctacatctgagataaactcaactcgtcgtcgaagaagactgagaagaagaag 682
 DB 3545 cacatgataatcttgagataaataatcttaactccttctgtaaggaacttaaaatgaaagaag 3604
 QY 683 aatcaaaagacgtcgcagacaacagctccaactcgtgtagcctcgtgaagaactcctcgggtg 742
 DB 3605 aaatcaaaagattatagatcaatcaatcaatcaggtattttaaagaactcttgggggtg 3664
 QY 743 actaactgcgtgtagacaacacgctactacaatgctgaatcgtcgtacagatccgaacaatacg 802
 DB 3665 attatttacaatatgataaaccatactataatgtaattatataatgataccaataataatg 3724
 QY 803 ttgacgtcaacaatgtaggtatccgcgtgtacatgtagcgaagaagcgcggtggtcgtg 862
 DB 3725 tctgtagtaaatatgtaggtatcagaggtatataatgtaactttaaaggccagagagtagcg 3784
 QY 863 ttatgactacaacatctactcgtgaactcctcctgtagcgtggtgtagcacaatctcatca 922
 DB 3785 taatgactacaacatcttatttaattcaagttagtagggggaacaataattattataa 3844
 QY 923 agaaatacgcgtctgtagaacaagaacataatgctgcgaacaatgtagtatactca 982
 DB 3845 aaaaataatgcttctgtaaaataaagataatctgtagaataatgtagcgtgtagtata 3904
 QY 983 atgtttagttagaagaagaatagatcgtcgtcgaataatgcttctcagcgtgtagt 1042
 DB 3905 atgttagttagtaaaataaagaatataaggttagctactaactgatacagcagcgtag 3964
 QY 1043 aagaagatctgtagtctgtagaataccgcagcgtgtagaactcgtcctcaggtgtagt 1102
 DB 3965 aaaaaatacgaagtgcatttagaataactcgtatgtagaatactcaatgtagcaggtgtagt 4024
 QY 1103 tgaatccagaagaacagcaggtatatacctaacaatgcaaatgtaactcgtcaggaacaa 1162

DB 4025 tgaagtcataaaatgatacgaagatacaaaataaatgcaaatgatttcaagaataata 4084
 QY 1163 atgtagacatcgtatcgttcatcgttccacaggtcaacaatcagtaactcgtg 1222
 DB 4085 atgggaatgataatagcttataagatttcaacagtttaataatagctcaactagtag 4144
 QY 1223 ctccacactgtagaactcagatcgaacgtctcctcctcgcactcgtggtgctggg 1282
 DB 4145 caagtaattgtagataatagacaatagaagaactagtagaacttgggtgctcatg 4204
 QY 1283 agttcaatccggtttagatgaacggttgggtgaagtcgcgtgtagaattc 1332
 DB 4205 aattattcctgtagatgtagatgtagtggggaagaagccactgtaataatc 4254

RESULT 13
 AAZ87220
 ID AAZ87220 standard; DNA: 1317 BP.
 XX
 AC AAZ87220;
 XX
 DT 08-MAY-2000 (first entry)
 XX
 DE DNA encoding native BONT serotype A (BONTA) C-terminal fragment (HC).
 XX
 KW Botulinum neurotoxin; heavy chain; BONT; serotype A;
 KW C-terminal fragment; HC; Venezuelan equine encephalitis virus replicon;
 KW VEE; botulism; vaccine; diagnosis; drug screening; ds.
 XX
 OS Clostridium botulinum.
 XX
 FH Key Location/Qualifiers
 FT 1..1317
 FT CDS /tag= a
 FT /product= "Native botulinum neurotoxin serotype A
 FT (BONTA) heavy chain C-terminal fragment"
 FT /transl_except= (pos:1294..1314, aa:Gly)
 XX
 PN MO200002524-A2.
 XX
 PD 20-JUN-2000.
 XX
 PF 09-JUL-1999; 99MO-US15570.
 XX
 PR 10-JUL-1998; 98US-0092416.
 PR 12-MAY-1999; 99US-0133870.
 XX
 PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
 XX
 PI Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
 XX
 DR WPI: 2000-160827/14.
 DR P-PSDB; AAY77142.
 XX
 PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
 PT toxin serotypes A-G, is used for inducing an immune response against
 PT botulinum -
 XX
 PS Example 3; Page 52; 54pp; English.
 XX
 CC The invention relates to novel vaccines that induce a protective immune
 CC response against botulinum neurotoxin (BONT) serotypes A, B, C, D, E, F
 CC and G (BONTA-BONTG). The vaccine of the invention is novel recombinant
 CC DNA construct comprising a vector, and at least one nucleic acid
 CC fragment comprising a C-terminal heavy chain fragment (HC) from BONT
 CC serotypes A-G. In preferred embodiments of the invention, the vector is a
 CC Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this
 CC vector results in the production of large amounts of a protein encoded by
 CC a sequence cloned into the replicon. The constructs are used to produce
 CC vaccines against botulism. The proteins can also be used as diagnostic
 CC tools for the diagnosis of botulism. The transformed host cells can be
 CC used to analyse the effectiveness of drugs and agents which inhibit toxin

CC effects. The vaccine currently used against botulism is dangerous
 CC and expensive to produce, and contains formalin, which is very painful
 CC for the recipient. Also, the vaccine is incomplete, in that only 5 of
 CC the 7 serotypes are represented in the formulation. The novel vaccine
 CC of overcomes these problems, as it is easily purified, and available in
 CC large quantities. It is also expressed in the lymph nodes for a better
 CC immune response. The present sequence represents DNA encoding native
 CC BONTA heavy chain C-terminal fragment (Hc) used in an exemplification of
 CC the present invention.

XX
 SQ Sequence 1317 BP; 560 A; 108 C; 211 G; 438 T; 0 other;

Query Match 54.5%; Score 726.4; DB 21; Length 1317;
 Best Local Similarity 72.3%; Pred. No. 2.8e-194;
 Matches 943; Conservative 0; Mismatches 361; Indels 0; Gaps 0;

QY 23 tgtctacctgaatacatcaagaacatcatcaccctccacgtgacgtcgct 82
 DB 14 tctctacattctgaatatttaagaatatttactcttctattggaatttaagat 73
 QY 83 acgaatccaatcactgtagcagcctgctcgtctgtagcgtcccaaaatcaacatcggttcta 142
 DB 74 atggaagtaactcatcttaaatagacctctcaggtatgcatcaaaaataaatattgtagta 133
 QY 143 aagttactcgcagtcgacgaagaatcagatccagctggttcaactcgcgaattctca 202
 DB 134 aagtaaatcttgatccaatagataaaatacaatcaattcaattatttaattagaagtagta 193
 QY 203 aaatcgaagtatcctggaagaatgctatcgatatacaactctatgtgaagaaacttcaca 262
 DB 194 aaattgaggttaattttaaanaaaatgctatgtatataatagatgtatgtaaaaatttagta 253
 QY 263 cctctctcgtgagtcgtagtccggaatacttcaactccactcctcctggaacaatgtaata 322
 DB 254 ctgagcttctggaagaatctcctaagatcttcaacgataagtcctaataatgtaata 313
 QY 323 ccatcatcaactgcatggaanaaactctggttggaaaatctctgaactacgtgtaaa 382
 DB 314 caataataattgtatggaanaaataatcagatggaagatcactcaattatgtgtaaa 373
 QY 383 tcaatctggaactcgcaggaactcaggaatacaaacagcgtgtgtatcaataactctc 442
 DB 374 taatctggaactcgcaggaactcaggaatacaaacagcgtgtgtgtttaaatacagtc 433
 QY 443 agatgatcaacatcctcgtgacatcaatcgcgtgagatcctggtaccacccaataatc 502
 DB 434 aaatgattcaatataatcagattatataaacaagatggttttgaactaccataataa 493
 QY 503 gtctgaataactccaanaatctacatcaacgcgctctgtagcagcagaacacgcatctca 562
 DB 494 gattaaataactcaanaatattatataaatggaagttaatagatcaaaaacaaattcaa 553
 QY 563 atctgggttaacatcaacgctcttaataacatcattcaactggaaggtgtcgtgaca 622
 DB 554 atttggttaatttctcagtagtaataataatgatttaattgaatggtgttagagata 613
 QY 623 ctccagcgtacatcctggaactcaactctcgtgtgtagcagaagaagcgaagaag 682
 DB 614 caatagatataattggaataaataatttcaatctcttctgtagaagattcaaatgaanaag 673
 QY 683 aaatcaagaacctgtagcagaacacagtcgaactctgtagtccgtagaagaactcgtgggtg 742
 DB 674 aaatcaagaacttataatgataatcaatcaatcaggtatctttaaagaagcttgggggtg 733
 QY 743 actacctgtagcagaacacccgtatctacatgctgaatctgtagcagtcggaacaataag 802
 DB 734 attattcaaatatgataaaccataactatattgtaattatattatgccaataataatag 793
 QY 803 ttgaagtaacaatgtagttagcgcggttatactgtagaagaagtcggtgtgtctg 862
 DB 794 tcgagtaataatgtagttagcgcggttatactgtagaagaagtcggtgtgtctg 853

QY 863 ttatgactacaacatctactcgaactcttcctcgtgtagtcccaatcatcatca 922
 DB 854 taatgactacaacatctattttaaattcaagtttgatagggggacaatttatata 913
 QY 923 agaaatcgcgtcgtgtaacaaggacaatatcgttcgcaacaatgctgtatataca 982
 DB 914 aaaaatctctcgtggaataaagaataatctgttagaataatgactgtatata 973
 QY 983 atgtgtagtagaacaagaataaccgctcgtcgtcgaactccttcaggtgtgtg 1042
 DB 974 atgttagtagttaaataaagaataataggttagtctactaattgcatcagcagcggtag 1033
 QY 1043 aaaagatctgtcgtcgtcgtgaatcccggaactgtgtgaatcgtgtcagtagttagta 1102
 DB 1034 aaaaatactaggttagttagaataatcccgatgttaggaatcgaagtagta 1093
 QY 1103 tgaatccaagaacgacaggggtatcactcaacaatgcaaatgaaatctgcgagaca 1162
 DB 1094 tgaagtcaaaataatgatacgaagaatacaataatgcaaaaatgaaatttcaagaata 1153
 QY 1163 atgttagatatacgtttcattcgttcacgaatcgaatcgaatcgttaactgtg 1222
 DB 1154 atgggaatgataaggtttagataggttccatcagtttaataatagctaaactagtag 1213
 QY 1223 ctcccaactggtacaaatcgtcagatcgaacggttcctcctcgcactcgtggtgtcgtgg 1282
 DB 1214 caagtaattggtataatagtagacaataatgaagaactagtagtgcgttgggtgtcgtatgg 1273
 QY 1283 agttaccccggttgatgacggttggtggtggaacgtcgcgtgtgaa 1326
 DB 1274 aatttattcctgtagatgtagtggggaagaagccactgtgaa 1317

RESULT 14
 AAZ87218
 ID AAZ87218 standard; DNA; 2532 BP.
 XX
 AC AAZ87218;
 XX
 DT 08-MAY-2000 (first entry)
 XX
 DE DNA encoding native botulinum neurotoxin serotype A (BONTA).
 XX
 KW Botulinum neurotoxin; heavy chain; BONT; serotype A;
 KW Venezuelan equine encephalitis virus replicon;
 KW VEE; botulism; vaccine; diagnosis; drug screening; ds.
 XX
 OS Clostridium botulinum.
 XX
 FH Key
 FT CDS 1..2532
 FT Location/Qualifiers
 FT /tag= a
 FT /product= "Native botulinum neurotoxin serotype A
 FT (BONTA) heavy chain"
 FT /transl_except= (pos:2509..2529, aa:Gly)
 PN MO200002524-A2.
 XX
 PD 20-JAN-2000.
 XX
 PF 09-JUL-1999; 99MO-US15570.
 XX
 PR 10-JUL-1998; 98US-0092416.
 PR 12-MAY-1999; 99US-0133870.
 XX
 PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
 XX
 PI Lee JS, Pushko P, Smith JF, Parker M, Dertbaugh MT, Smith LJ;
 DR WPT. 2000-160827/14.
 DR P-PSDB; AA77140.
 XX
 PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 17:11:08 ; Search time 139.08 Seconds
(without alignments)
2352.487 Million cell updates/sec

Title: US-09-611-419A-1

Perfect score: 1332
Sequence: 1 gaatcgaacagatcgtctt.....aacgtccgtgtaagaattc 1332

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1311	98.4	1330	1	US-08-480-604A-22
2	1311	98.4	1330	2	US-08-405-496A-22
3	1311	98.4	1330	4	US-08-915-136-22
4	1311	98.4	1402	1	US-08-480-604A-25
5	1311	98.4	1402	2	US-08-405-496A-25
6	1311	98.4	1402	4	US-08-915-136-25
7	726.4	54.5	3891	1	US-08-480-604A-27
8	726.4	54.5	3891	2	US-08-405-496A-27
9	726.4	54.5	3891	4	US-08-915-136-27
10	166.4	12.5	1359	1	US-07-618-312A-3
11	166.4	12.5	1359	1	US-08-280-228-3
12	74.2	5.6	1359	1	US-07-618-312A-1
13	74.2	5.6	1359	1	US-08-110-786A-7
14	74.2	5.6	1359	1	US-08-280-228-1
15	1858	5.6	1858	1	US-08-668-381A-6
16	60	4.5	5511	3	US-08-928-361B-2
17	60	4.5	7334	3	US-08-928-361B-1
18	57	4.3	5163	3	US-08-700-651-1
19	57	4.3	5163	3	US-08-928-361B-4
20	57	4.3	5318	3	US-08-700-651-2
21	57	4.3	5318	3	US-08-928-361B-3
22	40	3.0	1690	1	US-08-276-452A-24
23	40	3.0	1690	2	US-08-798-744-24
24	37.8	2.8	2277	1	US-08-676-967-2
25	37.8	2.8	2277	1	US-08-676-974-2
26	37.8	2.8	2277	2	US-09-098-487-2
27	37.2	2.8	18596	4	US-09-318-448-11

28	36.8	2.8	1430	1	US-08-276-452A-25	Sequence 25, Appl
29	36.8	2.8	1430	2	US-08-798-744-25	Sequence 25, Appl
30	34.8	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
31	34.6	2.6	2094	3	US-08-714-918-87	Sequence 87, Appl
32	34.6	2.6	2094	4	US-09-265-315-87	Sequence 87, Appl
33	34.6	2.6	2094	4	US-09-265-315-87	Sequence 87, Appl
34	34.6	2.6	2094	4	US-09-265-315-87	Sequence 87, Appl
35	34.4	2.6	951	1	US-07-855-412B-2	Sequence 87, Appl
36	34.4	2.6	951	2	US-08-308-887A-2	Sequence 2, Appl
37	34.4	2.6	951	3	US-08-881-094-12	Sequence 2, Appl
38	34.2	2.6	198	5	PCT-US95-10668-1	Sequence 1, Appl
39	34.2	2.6	198	5	PCT-US95-10668-2	Sequence 1, Appl
40	34.2	2.6	198	5	PCT-US95-10668-3	Sequence 3, Appl
41	34.2	2.6	198	5	PCT-US95-10668-4	Sequence 4, Appl
42	34.2	2.6	3292	1	US-07-814-964-12	Sequence 12, Appl
43	34.2	2.6	3292	1	US-08-258-442-12	Sequence 12, Appl
44	34.2	2.6	3292	1	US-08-328-809-7	Sequence 7, Appl
45	34.2	2.6	3292	5	PCT-US92-11107-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-22
; Sequence 22, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCHA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLTA, DIANE E.
; REGISTRATION NUMBER: 40,027

```

REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1314
US-08-480-604A-22

```

```

Query Match      98.4%; Score 1311; DB 1; Length 1330;
Best Local Similarity 100.0%; Pred.No.0;
Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 16 cgtctgtcttaaccttaactgaatacatcaagaacatcatcaatcatcctgaac 75
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 7 cgtctgtcttaaccttaactgaatacatcaagaacatcatcaatcatcctgaac 66
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 76 ctgagctagaatccaatccaatcctgagctctgctgaagctcccaaatcaatc 135
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 67 ctggcgctacgaatccaatccaatcctgagctctgctgaagctcccaaatcaatc 126
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 136 gttcttaagttaaccttaactgaatacatcaagaacatcatcaatcatcctgaac 195
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 127 ggttttaagttaaccttaactgaatacatcaagaacatcatcaatcatcctgaac 186
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 196 tcttccaataatcgaaatgatactcgaagaatgatactcgaagaatgatactcgaagaac 255
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 187 tcttccaataatcgaaatgatactcgaagaatgatactcgaagaatgatactcgaagaac 246
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 256 tcttccaatcctctctgagctgaatcgaagaatgatactcgaagaatgatactcgaagaac 315
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 247 tcttccaatcctctctgagctgaatcgaagaatgatactcgaagaatgatactcgaagaac 306
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 316 gaatacaccaatcaactgaatgaagaagaatctggttgaagaatgatactcgaagaac 375
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 307 gaatacaccaatcaactgaatgaagaagaatctggttgaagaatgatactcgaagaac 366
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 376 ggtgaataatcctgagctgaatgaagaagaatctggttgaagaatgatactcgaagaac 435
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 367 ggtgaataatcctgagctgaatgaagaagaatctggttgaagaatgatactcgaagaac 426
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 436 tactctgaatgatactgaatcctgagctgaatgaagaagaatctggttgaagaac 495
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 427 tactctgaatgatactgaatcctgagctgaatgaagaagaatctggttgaagaac 486
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 496 aacaaatgctgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 555
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 487 aacaaatgctgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 546
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 556 atctccaatctggtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 615
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 547 atctccaatctggtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 606
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 616 ctggaacatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 675
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 607 ctggaacatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 666
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 676 gaaataaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 735
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 667 gaaataaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 726
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 736 tgggtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 795
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 727 tgggtgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 786
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

QY 796 aaatacgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 855
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 787 aaatacgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 846
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 856 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 915
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 847 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 906
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 916 atcaataaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 975
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 907 atcaataaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 966
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 976 taatacgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1035
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 967 taatacgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1026
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1036 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1095
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1027 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1086
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1096 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1155
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1087 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1146
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1156 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1215
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1147 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1206
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1216 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1275
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1207 gttctgttaacgttaacaaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1266
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1276 tcttgaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1326
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1267 tcttgaagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaac 1317
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 2
US-08-405-496A-22
Sequence 22, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
NUMBER OF INVENTIONS: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791

```


	Query Match	98.4%	Score 1311	DB 2	Length 1402
	Best Local Similarity	100.0%	Pred. No. 0		
	Matches 1311	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0
Oy	16	cgctgcgtctcactcttcacgtacatcacagaacatcatcaatccctcaac	75		
Db	79	cctctgcgtctcactcttcacgtacatcacagaacatcatcaatccctcaac	138		
Oy	76	ctgcgtacgaatccaatcaatcgtacgactgtcgtctacgcttccaaatcaacatc	135		
Db	139	ctgcctcctcgaatccaatcactcgtacgactgtcgtctacgcttccaaatcaacatc	198		
Oy	136	ggtctaaagttaacttcgatcgcgatcgcagaagaatcgaatccagctgttcaatctgga	195		
Db	139	ggttcttaagtttaatttgcattccgattccgaaatcgaatcgaatccagctgttcaatctgga	258		
Oy	136	tcttccaaaatcgaagtatacctgaagaatgctatcgtatatacaactctatgtacgaanaac	255		
Db	259	tcttccaaaatcgaagtatacctgaagaatgctatcgtatatacaactctatgtacgaanaac	318		

OY	256	ttcccaaccccttcttgagatccgatccggaataacttcaactcattcttcgaaacat	315
Db	319	tttctccacccctcttcttgagatccgatccggaataacttcaactcattcttcgaaacat	378
OY	316	gaataacacatcatcaactctcattggaanaaacaatttcggtttgnaagta tctctgaactac	375
Db	379	gaataacacacatcatcaactctcattggaanaaacaatttcggtttgnaagta tctctgaactac	438
OY	376	gggtgaatcatctcygactctcygagcaactcaaggaaatcaaacagcgtgtgtattcaaa	435
Db	439	gggtgaatcatctcygactctcygagcaactcaaggaaatcaaacagcgtgtgtattcaaa	498
OY	436	tactctgaatgatcaacaactctctgactcaatcaacacgcgtgactcttgtaacatacc	495
Db	499	tactctgaatgatcaacaactctctgactcaatcaacacgcgtgactcttgtaacatacc	558
OY	496	aacaatcgcgtctgaaataactccaaatctcatcaatacgcgcgtctgatacgcacgaacacg	555
Db	559	aacaatcgcgtctgaaataactccaaatctcatcaatacgcgcgcgtgatacgcacgaacacg	618
OY	556	attcccaatctgggttaacatccacgcgttttaatacatatgattcaaatctgaacggttgt	615
Db	619	attcccaatctgggttaacatccacgcgttttaatacatatgattcaaatctgaacggttgt	678
OY	616	cgttaacactcaacgctctacatcttgatcaaaataactcaatctcgtttcgcgaanaaactgaac	675
Db	679	cgttaacactcaacgctctacatcttgatcaaaataactcaatctcgtttcgcgaanaaactgaac	738
OY	676	gaaagaagaatcaagaacgcgtgtatgcgaacacagttccaaattcgtgtatccgtgaagaacttc	735
Db	739	gaaagaagaatcaagaacgcgtgtatgcgaacacagttccaaattcgtgtatccgtgaagaacttc	798
OY	736	ctggggttacctaccctcgcagtagcgaacaaacgcttaactatgctgaatctgtatcatccgcagc	795
Db	799	ctggggttacctaccctcgcagtagcgaacaaacgcttaactatgctgaatctgtgtacgattccgacac	858
OY	796	aaatacgtctgacgylcaacaaatgtaggtatccgcggtttacatgyltaccctgaaaggtccgcgt	855
Db	859	aaatacgtctgacgylcaacaaatgtaggtatccgcggtttacatgyltaccctgaaaggtccgcgt	918
OY	856	ggttcttgtttagatctaccacaactctcaactgaactcttcctgttaccggtggttacaacatc	915
Db	919	ggttcttgtttagatctaccacaactctcaactgaactcttcctgttaccggtggttacaacatc	978
OY	916	atcatacaagaataatccggtctcgttgaacaaaggaacaatatcgttcgcgaanaatgatacgttgta	975
Db	979	atcatacaagaataatccggtctcgttgaacaaaggaacaatatcgttcgcgaanaatgatacgttgta	1038
OY	976	tacatcaaatgttgttagttaagaacaagaataacacgtctcgtgtcacaacatgctttccaagct	1035
Db	1039	tacatcaaatgttgttagttaagaacaagaataacacgtctcgtgtcacaacatgctttccaagct	1098
OY	1036	gggttgaagaagaactctcgtctcgtctcgtggaataatccgcgaacgtttggtcaatcgtctcaagta	1095
Db	1099	gggttgaagaagaagaactctcgtctcgtctcgtggaataatccgcgaacgtttggtcaatcgtctcaagta	1158
OY	1096	gctgtaatgaaatcccaagaacgcgacgggtatcatcactaacaatgycanaatgaaatcgtcag	1155
Db	1159	gctgtaatgaaatcccaagaacgcgacgggtatcatcactaacaatgycanaatgaaatcgtcag	1218
OY	1156	gacacaatgtaacggtatcgggtttcaatcgggtttccaccacgattcaaaatatcgtctaa	1215
Db	1219	gacacaatgtaacggtatcgggtttcaatcgggtttccaccacgattcaaaatatcgtctaa	1278
OY	1216	ctgggtccttcccaacgtgtatacaatcgtcagaatcgaacgcttccctcctcgcgaactctgggttgc	1275
Db	1279	ctgggtccttcccaacgtgtatacaatcgtcagaatcgaacgcttccctcctcgcgaactctgggttgc	1338
OY	1276	tctctgggaatcatcccggtttgattgacggtttgggttgaacgtccgcgtgttaa	1326
Db	1339	tctctgggaatcatcccggtttgattgacggtttgggttgaacgtccgcgtgttaa	1389


```

RESULT 6
US-08-915-136-25
: Sequence 25, Application US/08915136
: Patent No. 6290960
: GENERAL INFORMATION:
: APPLICANT: KINK, JOHN A.
: APPLICANT: THALLEY, BRUCE S.
: APPLICANT: PADHAY, NISHA V.
: APPLICANT: FIRCA, JOSEPH R.
: APPLICANT: STAPFORD, DOUGLAS C.
: TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
: TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL, LLP
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/915,136
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/480,604
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/405,496
: FILING DATE: 16-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/329,154
: FILING DATE: 25-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,907
: FILING DATE: 02-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/985,321
: FILING DATE: 04-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/429,791
: FILING DATE: 31-OCT-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: INGOLIA, DIANE E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: OPHD-01763
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1402 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1386
: US-08-915-136-25

```

```

Query Match          98.4%; Score 1311; DB 4; Length 1402;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 16 cgtctgcttaactcaactgaataacatcaagaacatcaataactcctgaac 75
|||||

```

```

Db 79 CGTCTGCTGTCTACTTCACTGATCAATCAAGAACATCAATCACTCCATCCGTAAC 138
QY 76 ctggcgtaagaatcaataacttgatcgccgtgtcgttaacttccaataactc 135
Db 139 CTGCGCTACGAATCCAAATCACTGATCGACGCTGCTGCTACGCTTCCAAAATCAACATC 198
QY 136 ggtcttaagtttaacttcgatccgatcgacaagaatcaatcaactcagctgttcaactcgtga 195
Db 199 GGTTCGAAAGTTAACTTCGATCCGATCGACGACGAAGATCAATCACTCAGCTGTTCATCTCGGA 258
QY 196 tcttccaataatcgaagtatccttgaagaatgtctatcgtatataactctatgtagaanaac 255
Db 259 TCTTCCAAAATGGAAGTTATCTGTAAGAAATGCTATGATACACTATATGATCAAGAAAC 318
QY 256 tttctcactctcttctgagatccgtatccgaaatcttaactcaatctcttgaacat 315
Db 319 TTCTCCACTCTCTTCTGGATCCGTATCCGGAATATCTTAACCTCCATCTCTGGAACAT 378
QY 316 gaatacaccatcatcaactgcatacgaagaacattctgttgaagaatctctgaactac 375
Db 379 GAATACACCATCATCAACTGATGGAAGAAACAAATCTGTTGGAAGTATCTGGAACATC 438
QY 376 ggtgaataatcatctggaactctcagaagacctcaggaatacaacagcgtgtgtatcaaa 435
Db 439 GGTGAATATCATCTGCACTCTGCAAGCACTCAAGAAATCAACAGCGTGTGATTCANA 498
QY 436 tactctcagatgatcaaatctctgtactatcaatcagcttgatcttctgaactacc 495
Db 499 TACTCTCAGATGATCAATCTCTGACTACATCAATCGTGGATCTTTCCTTACCATCAC 558
QY 496 aacaatcgtctgaataactccaataatctcaatcaacagcgctgtgatacgaagaacgg 555
Db 559 AACAAATCGTGTGAATATCACTCCAAAATCTATCAACAGCGCTGTGATGACCAAGAAACGG 618
QY 556 atctccaatctggttaacaatccaagccttcaataacatgatttcaactgtagcgtgt 615
Db 619 ATCTCCAATCTGGGTAACTCACTCCAGCGCTCTTAATATCATCATGTCAAACTGAGCGGTGT 678
QY 616 cgtgaactcaccggtacacttgatcaaatcttcaactcttctgaagaagaactgaac 675
Db 679 CGTGACACTACCGGCTACATCTGATCAATCTTAATCTGTGCACAAAGAACGTAAC 738
QY 676 gaaaaagaatacaagacctgtacgaacaacagctccaattctgttaccgtgaagaactc 735
Db 739 GAAAGAAATATCAAGACCTGTACGACACACAGTCCAAATTCGTGATCCGTGAAGACTTC 798
QY 736 tgggtgacttaactcagtagaacaacacgctactacatgctgaatctgtacatccgaac 795
Db 799 TGGGGTGACTACTGCACTGACGACAAACCGTACTACATGCTGAATCTGTATGACATCCGAAC 858
QY 796 aaatcgttgaggtcaacaatgtagtatccggtgtatcatgtacctgaagaagctccggt 855
Db 859 AAATACGTTGACGTCAACATGATGATCCGCGGTATCATGTACTGGAAGGTCGCGT 918
QY 856 ggtctgtatgactacccaacatctcactgaactcttccctgtacccgtgtgtacccaattc 915
Db 919 GGTTCGTATGACTACCAACATCTACCTGCAACTCTTCCCTGTACCGTGTGACCAATATC 978
QY 916 atcatcaagaataacgcgtctgtgtlaacaagaacaalatgcttcgacaacatgactgtgta 975
Db 979 ATCATCAAGAATAATACGCGCTGTGTAACAAGGACAAATATGTTGCGCAACATGATCGTGA 1038
QY 976 tacatcaatggtgtgtagtaagaacaagaatatccgtctgtctacaatgtcttcaaggt 1035
Db 1039 TACATCAATGTTGATTAGTAAGAACAAAGATACCGCTGTGCTACCAATGCTTCTCAAGGCT 1098
QY 1036 ggttagaagaagatctgtctgtctcgtgaatcccggaacgctgtgtaactgtctcaggt 1095
Db 1099 GGTGTAGAAATAATCTGTCTGTCTCTGGAATTCGCGAGAGTGTGTATCTGTCTCAAGTGA 1158
QY 1096 gttgttaatgaatccaagaacgacagggatcatcaataatgcgaataatgactcag 1155
Db 1159 GTTGTAAATGAATTCACAGAACGACGAGGTATACATAAATGCAAAATGAAATCTGCAG 1218

```


Db 3488 AAAAATGCTTCTGAAAAAATAAATGTTAGAAATATGATCGGTATATATTA 3547
Qy 983 atgttgatgaagaacaaagataccgtgctacccatgcttccagcgtggtag 1042
Db 3548 ATGTAGTAGTTAAATAAATAAGATATAGGTAGTACTTAATGATACAGCGGCTAG 3607
Qy 1043 aaaaagatctgtctgtctgtgaatcccgagcgttgtaatccgtctcaggtagttaa 1102
Db 3608 AAAAATAGTAACTAGCATTAAGAAATACCTGATGTAGAAATCTAAGTCAAGTAGTA 3667
Qy 1103 tgaatccagaagaacccaggtatcactaaatgcaaatgaaatctgcaggaacaa 1162
Db 3668 TGAAGTCAAAAATGATCAAGGAATACAAATTAATGCAAAATGAAATTCACAAATTAATA 3727
Qy 1163 atgttgatgatatcggttctcatcggttccacagttcaacaatctcgtcaaatcgtgtg 1222
Db 3728 ATGGGAATGATATAGGCTTATAGATTCATTCATGTTTAAATATAGCTAAACTAGTAG 3787
Qy 1223 ctcccaactggtacaaatcgatcgaaacgcttcctcctcgactcgtggtgtctgtg 1282
Db 3788 CAAATGATGTTATATAGCAATATAGAAATCTAGTAGGACTTGGGCTCATGGC 3847
Qy 1283 agtcaatcccggttgatgacggttggtggtgacgctcgtgtgtaa 1326
Db 3848 AATTATTCCTGTAGATGATGATGGGAGAAAGCCACTGTAA 3891

RESULT 8
US-08-405-496A-27
; Sequence 27, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3888
US-08-405-496A-27

Query Match 54.5%; Score 726.4; DB 2; Length 3891;
Best Local Similarity 72.3%; Pred. No. 7.9e-204;
Matches 943; Conservative 0; Mismatches 361; Indels 0; Gaps 0;

Qy 23 tgtctactctactgaatatacataagaacatcaatactccatccgaactcgtgcgt 82
Db 2588 TATCTACATTTACTGTAATATATTAAGAAATATATTAATCTTCTTAATTTGAATTAAGAT 2647
Qy 83 acgaatccaaatccatcgatcgacctgtctgcgaactcccaaatcaaatcgtgtcta 142
Db 2648 ATGAAGTAATATCAATTAATAGACTTATCTAGTATGATGATCAAAAATTAATTTGTA 2707
Qy 143 aagttactcgatccgatcgacaagaatcagatccagctgttcaactcgtgaacttcca 202
Db 2708 AAGTAAATTTTGAATCCATATGATTAATAATCAATTCATATTTTAATTTGAAGTAGTA 2767
Qy 203 aaatcgaaagttatccctgaagaatgctatcgatatacaactctatcgaaacttcca 262
Db 2768 AAATGAGTAATTTTAAATAAATGCTATGTATATATAGTATAGTAAATTTTAGTA 2827
Qy 263 cctcctcttgatcgatccgaatcactccatccatctctcgtgaactgaatgaataca 322
Db 2828 CTAGCTTTGGATTAAGATTCCTAGATATTTTAACAGTATAGCTTAATTAATTAATTA 2887
Qy 323 ccatactcaactcgtatgaaacaatctcgttggaagaatctcgtgaactcgtgaactga 382
Db 2888 CAATTAATTAATTTGATGGAATAATTAATTCAGATGGAAGATATACCTTAATTTAGTAGTA 2947
Qy 383 tcaatcgacctcgtcaggaacatcagaagaatcaacaacgctgtgtatcaactctc 442
Db 2948 TAATCTGAGACTTACAGGATCTCAGGAATAAACAAGATGTTTAAATACAGTTC 3007
Qy 443 agatgatcaaatctcgtatcatatcaatcgcgtgacatcgttcatatcacaacaatc 502
Db 3008 AAATGATTAATTAATATGATTAATTAACAGATGATTTTGTAACTATCTAATTAATA 3067
Qy 503 gtcgataaactccaaatctatatacaacgctcgtatcgacacgaacacgatctcca 562
Db 3068 GATTAAATATCTCTAAATTTATATAAATGGAAGATTAATGATCAAAAACCAATTTCA 3127
Qy 563 atctgggtaaatccacgcttctaataacatcagttcaaatcgtgacggtgtcgtgaca 622
Db 3128 ATTTAGTAATTAATTCATGCTAGTAATTAATTAATGTTTAAATTAATGATGTTGTA 3187
Qy 623 ctacccgctacatcgtatcaaatcactcaatctgttcgcgaacacgtgaacgaagaag 682
Db 3188 CACATAGATATATTTGATTAATAATTTTAACTTTTGTAAAGATTAATTAAGAAAAG 3247
Qy 683 aaatcaagaacctgtacgacaacacagttccaatctggtatccgtgaagaactctgggtg 742
Db 3248 AAATCAAGATTTATATGATTAATCAATCAATCAATCAATCAATCAATCAATCAATCA 3307
Qy 743 actaactcgatcgacaacacgcttactatcgttgaatctgttgcgtacgacaacatag 802
Db 3308 ATTTATTTCAATATGATTAACCAATCAATCAATCAATCAATCAATCAATCAATCAATG 3367
Qy 803 ttgaactcaacaatgtaggtatcgcggtgtacgtacacgtgaagaactcgcggtgtcgt 862
Db 3368 TCATGTAATTAATATGATGATTAAGAGTTATATGATTTTAAAGGCCCTAGAGTAGCG 3427
Qy 863 ttatgactacaacatctactcgaactctccctgatacgttggtlaccaaaatcatcata 922

Oy	683	aacaaagagccgtgtagacaacagtcacatctcttgatctcctgaaagactctggggtg	742
Db	3248	AAATCAAGATTATATAGATAAATCAAAATTCAGATATTTTAAACACTTTGGGGTG	3307
Oy	743	actaccgcagtagacaacacg tactacatgcgaatctgacatctgcagacacgaatacg	802
Db	3308	ATTATTTTACAATATGATAAACCATACTATATGTTAATTTATATGATCCAAATAAATATG	3367
Oy	803	ttgacgtcaacaatgtaglataccgcggtlatactgacatgaaaggtccgcgtgltctg	862
Db	3368	TCGATGTAAATTAATGTAGATATAGAGGTTAATATGATCTTAAAGGGCTGTAGAGTAGCG	3427
Oy	863	ttatgataccaacactctacctcgaactctccctccacgcgtgtgacaaatcaatca	922
Db	3428	TAAATGACTACAAACATTTATTTAAATTCAGATTTGTATATAGGGGCAAAATTTATTTAA	3487
Oy	923	agaaatacgcgtctggtlaaacaagaacataatcgltgcacaacatgatcgtlatacatca	982
Db	3488	AAAAATATGCTTCTCGCAATAAAGATAAATATTTGTGAATAATGATGCTGATATATTA	3547
Oy	983	atgtgtgaagttaagaacaaagaataccgcgtctggtctacaatgtcttcaagctgtgtag	1042
Db	3548	ATGTAGTAGTTAAAAATAAAGAAATATATAGGTTATAGCTATATCATCAACAGGACGGCTAG	3607
Oy	1043	aaaagatctgtctgtcctggaatcccgagcgttgtagatactgtctcagtgagtgttaa	1102
Db	3608	AAAAATATCTAAGTCGATTGGAATATACGTGATGTAGGAAATCTAAGTCAAGTAGTAGTAA	3667
Oy	1103	tgaataccaaagacgaccaggtgatactaactaacaatatgcaaaatgaactctgcagaaca	1162
Db	3668	TGAAGTCAAAAAATGATCAAGAAATATACAAATAAATATGCAAAATGATTTACAAAGATATA	3727
Oy	1163	atggttaacgatactcgggttcaatcaggtcttcacacagttcaacaatactgcctaacggtg	1222
Db	3728	ATGGGAATGATATATAGGCTTTATATAGGATTTCAATCAGTTTATATATATAGCTAACTAGTAG	3787
Oy	1223	cttccaactgtgtcaacatcgtcagatctgaaacgttctcctgcacatctgggttgcctctgg	1282
Db	3788	CAATATATTTGGTATTAATATAGCAAAATAGAAAGATCTATGTAGACCTTTGGGTTGCTCATGG	3847
Oy	1283	agttcaatcccggtgtgtagaagcgtgttggtgtgtaacgttccgctgtaa	1326
Db	3848	AATTTATTTCTGTATGATATGATATGGGAGAAAGGCACTGTGA 3891	

RESULT 10
US-07-618-312A-3
Sequence 3, Application US/07618312A
Patent No. 5389540
GENERAL INFORMATION:
APPLICANT: Makoff Dr, Andrew J
APPLICANT: Romanos Dr, Michael A
APPLICANT: Clare Dr, Jeffrey J
APPLICANT: Fairweather Dr, Neil F
TITLE OF INVENTION: VACCINES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: 14th floor
STREET: 2200 Clarendon Boulevard,
CITY: Arlington,
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22201
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/618,312A
FILING DATE: 19910516
CLASSIFICATION: 424

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER:  GB 8926832.0
3      FILING DATE: 28-NOV-1989
4
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER:  GB 9006097.1
7      FILING DATE: 17-MAR-1990
8      ATTORNEY/AGENT INFORMATION:
9      NAME: Crawford Mr, Arthur R
10     REGISTRATION NUMBER: 25,327
11     REFERENCE/DOCKET NUMBER: 510-51
12     TELECOMMUNICATION INFORMATION:
13     TELEPHONE: 0101 703 8750400
14     TELEFAX: 0101 703 5253468
15     TELEX: 200797 NIXN UR
16
17     INFORMATION FOR SEQ ID NO: 3:
18     SEQUENCE CHARACTERISTICS:
19     LENGTH: 1359 base pairs
20     TYPE: NUCLEIC ACID
21     STRANDEDNESS: double
22     TOPOLOGY: linear
23
24     MOLECULE TYPE: CDNA
25     HYPOTHETICAL: NO
26     ANTI-SENSE: NO
27     ORIGINAL SOURCE:
28     ORGANISM: Clostridium tetani
29     FEATURE:
30     NAME/KEY: CDS
31     LOCATION: 1..1356
32
33     US-07-618-312A-3

```

Query Match	Similarity	12.5%	Score	16.4%	DB 1:	Length	1359;
Best Local	Similarity	52.3%	Pred.	No. 2,9e-39			
Matches	486;	Conservative	0;	Mismatches	411;	Indels	33;
						Gaps	
QY	49	aacatcatcaatcaatccatccatccatgaacactcgactaagaatccaatccaatccatccatgaacactg	108				
Db	49	ATCCGTAAGAAAAAGTCAACCTTCGTGAACCTGGACATCAACAGAGATATATATCCCGCATAC	108				
QY	109	tctgcctcgcgtccatcaaaaatcaacactcggtcttaagttaacttgatccg---atgcac	165				
Db	109	TCTGCTTCACCTCCCTCTGTTATCAACATATCCAGATGCTCAATTTGGTGGCGGCGCATCAAC	168				
QY	166	aagaaatcagatccagctgttccaatctggaatctccaataatgaaagttaacctgaagaat	225				
Db	169	GGCAAGAGTATCCACCTGGTTAAACAAGCATCTTGTGAAGTTATGCTGCACAAGGCCATAG	228				
QY	226	gctatctgtataaactctatgttagaagaaactctccaactctcttgatctcgatcccg	285				
Db	229	GACATCGAATTAACAGACATGTTCAACAACCTTACCCGTAGTGGTTCGGCGCGCTGCCG	288				
QY	286	aaataactcaatcatc-----tctctgaacaatgaatatacaatcaatcaactgc	336				
Db	289	AAAGTTTTCGTGCTCCOACCTGGACAGCATAGCGGACATTAAGAGTACTCCATCAAGCTCT	348				
QY	337	atggaanaacat-----tctggttggaaagtctctctgaactaagtgtaa	381				
Db	349	ATGAAGAAAAACATCCCTGTCCATCGGCTCTGGTTGGTCTGTTTCCCTGAGAGGATACAAAC	408				
QY	382	atcatctggaactctgaagacaactcaagaaatcaaacagcgtgtgattacaataactct	441				
Db	409	CTGATCTGGACCTCTGAAGACATCCCGCGGCGCAAGTTTCGTCATCATCACTTCCGCGACCG	468				
QY	442	cagatgatcaacatctctgaactaacaatcagctggaatcttcgtttacaatccaacaat	501				
Db	469	CCGGACAAGTTTCAAGCGCGTACCTGGCTTAACAATAGGTTTTCATCATCACTTCACTTAACGAT	528				
QY	502	gcttcgaataactccaataatctacatcaaacagcgcttcgatcaccagaacacgattcc	561				
Db	529	CGTCTGTCTTCTGCTTAACCTGTACATCAACGGCGCTTCGATGGGCTCCGCTCAAAATCACT	588				
QY	562	aattcgggttaacatcaacgcttctaataacaatcaatgattccaacttgaacgtgttcgtgac	621				

Db 829 GTTACGCTGAAAA-----CATCAGCTACATGATGACACGCGCGCTCTAC 882
Qy 862 gtatgactccaacattaccctgaactctccctgtaacggtgtaccaatcatc 921
Db 883 ACTAAGGTAACATGACATCTACTACGACGCTGTGTAACAGCGCTTAATTCATCATC 942
Qy 922 aagaatacgcgtctgtaacgaagaacat 951
Db 943 AAACGCTACACTCCGACACGAAATCGAT 972

RESULT 12
US-07-618-312A-1
; Sequence 1, Application US/07618312A
; Patent No. 5389540
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 14th Floor
; STREET: 2200 Clarendon Boulevard,
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/618.312A
; FILING DATE: 19910516
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crawford Mr, Arthur R
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 510-51
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 0101 703 8750400
; TELEFAX: 0101 703 5253468
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-07-618-312A-1

Query Match 5.6%; Score 74.2; DB 1; Length 1359;
Best Local Similarity 47.3%; Pred. No. 3.9e-12;
Matches 223; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

Qy 349 tctgttggaagatctctctgaactcagtggaatcattctgactctgcagacactcag 408
Db 376 TCTGTTGGAGGTATCTACTTAAGGTAACTTAATATGAGCTTTAAAGATTCCGCG 435
Qy 409 gaaatcaaacagcgtgtgttattcaaatctctcagatgataacatctgactacac 468
Db 436 GGAGAGGTAGCAAAATTAATCTTTAGGCAATTTACCTGATTAATTAATGCTTTATTAGCA 495
Qy 469 aatcgtgattctgttaccatccacaacatcgtcgaataactccaatctacac 528
Db 496 AATTAATGGGTTTATTACTATTACTAATGATTAATATCTTCGCTAATTTGTATATA 555
Qy 529 aacgacctgctgacgaagaacagatctcaactctggtgaatactccacgttcaat 588
Db 556 AATGGAGTACTATTGGAAGTGCAGAAATTTACTGGTTTAGAGCTATTGAGAGATAT 615
Qy 589 aacatcatttcaactcagcgtgtgtctgtaactcaccgtcaactctgattcaatc 648
Db 616 AATATAACATTAATAACTAGATGATTAATTAATTAATCAATACGTTTCTATTGTATAA 675
Qy 649 tccaatctgttgcagaagaactcgaagaagaagaatcaagaactgtgacgaacacag 708
Db 676 TTTAGGATATTTTGCAGACATTAATCCAAAGAGATTGAAATTAATATACAGATTAT 735
Qy 709 tccaatctgttgcagaagaactcgtggtgtaactcgtgacgtgacgaacacgtac 768
Db 736 TTATCTATTAACCTTTTAAAGACATCTCTGGGAAACCCCTTACGATATGATACAGATAT 795
Qy 769 tacatgctgaatctgtacgataccgaacaaatcgttgaacgtcaacatgta 819
Db 796 TATTTAATACGATGCTTCTAGTTCTTAAGATGTTCAATGAAATATATA 846

RESULT 13
US-08-110-786A-7
; Sequence 7, Application US/08110786A
; Patent No. 5443966
; GENERAL INFORMATION:
; APPLICANT: FAIRWEATHER, Neil Fraser
; APPLICANT: MAKOFF, Andrew Joseph
; TITLE OF INVENTION: Expression of tetanus toxin fragment C
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye P.C.
; STREET: 1100 No. 5443966th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110.786A
; FILING DATE: 23-AUG-1993 1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/777.337
; FILING DATE: 29-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB90/00943
; FILING DATE: 20-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8914122.0
; FILING DATE: 20 June 1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary J. Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-134
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000

.....

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen, Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 17:08:43 : Search time 4365.92 Seconds
(without alignments)
4117.791 Million cell updates/sec

Title: US-09-611-419A-1
Perfect score: 1332
Sequence: 1 gaattcgaacagctgcgtct.....aacgtcgcgtctaagaattc 1332

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_cslba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_hiv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.4	4.4	608	10	BI887904 ZF637-1-0
2	54.6	4.1	1101	12	CNS00LO0
3	53.6	4.0	537	10	BM277996 AS_tgz-51
4	53.6	4.0	551	10	BM278689 AS_tgz-66
5	53.2	4.0	394	9	AU060224
6	52.6	3.9	447	10	BI594803 AS_tgz-44
7	52	3.9	597	10	BM278181 AS_tgz-54
8	51.8	3.9	500	10	BM278805 AS_tgz-67
9	50.6	3.8	511	10	BM004465 TGESTZya8
10	50.4	3.8	569	10	BM278657 AS_tgz-65
11	50	3.8	641	10	BM181884 FV51B1.Y
12	49.6	3.7	645	9	AI389106 GH20192.5
13	49.2	3.7	827	12	CNS02156
14	49	3.7	481	10	BF050030 AS_tgz-21
15	48.8	3.7	628	9	AU060230
16	48.2	3.6	546	10	C22974
17	48.2	3.6	1101	12	CNS0100X

18	48	3.6	546	10	BF169335
19	48	3.6	556	10	BM277853 AS_tgz-49
20	47.4	3.6	377	10	BI594814 AS_tgz-44
21	47.4	3.6	450	9	AU060996
22	47	3.5	458	10	BI783045
23	47	3.5	518	10	BI782765
24	46.8	3.5	1175	10	BI872945
25	46.4	3.5	469	10	BF050033
26	45.4	3.4	386	10	BF050064
27	45.4	3.4	499	10	BM278692
28	45.4	3.4	536	10	BF050073
29	45.4	3.4	542	10	BM278786
30	45.4	3.4	554	10	C23753
31	45.2	3.4	501	10	BM277916
32	45.2	3.4	525	12	BM184460
33	45.2	3.4	525	12	CNS07PEA
34	45.2	3.4	1101	12	CNS000D1
35	45	3.4	564	10	BE606733
36	44.8	3.4	534	10	BM278087
37	44.6	3.3	1101	12	CNS016JY
38	44.4	3.3	522	10	BM278558
39	44.4	3.3	681	12	CNS02EOD
40	44.2	3.3	325	12	CNS03PK8
41	44.2	3.3	429	9	AU052930
42	44.2	3.3	500	12	BE7199
43	44.2	3.3	519	10	BM278164
44	43.8	3.3	535	10	BI594948
45	43.6	3.3	443	10	BM039802

ALIGNMENTS

RESULT 1

LOCUS

BI887904 ZF637-1-002159 zebrafish shield stage whole embryo cDNA library

DEFINITION MPMGP637 Danio rerio cDNA clone MPMGP637_10F4:MPMGP637F0410 5', mRNA sequence.

ACCESSION BI887904 GI:16095175

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2001)
Contact: Hennig S
Laboratory 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
5' EST sequencing of clones from a zebrafish shield stage library,
normalised from 55,000 starting clones by oligonucleotide
fingerprinting
High quality sequence stop: 608.
Location/Qualifiers

FEATURES

SOURCE

1.608
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="MPMGP637_10F4:MPMGP637F0410"
/clone_lib="zebrafish shield stage whole embryo cDNA
library MPMGP637"
/tissue_type="whole embryo"
/dev_stage="shield stage, 6 hrs post-fertilisation"
/lab_host="E.coli, XLI blue MRF"

BASE COUNT	ORIGIN
207 a	218 c
41 g	140 t
2 others	2 others

Query Match	4.4%	Score 58.4	DB 10	Length 608
Best Local Similarity	49.1%	Pred. No. 0.00034		
Matches 181	Conservative 0	Mismatches 187	Indels 1	Gaps 1

Qy 448 atcaaacctctcgaactacatca - tccgtggaacttcgttacatcaccaacaatcgtct 506
Db 5 ATCAACTCTTCCAGCTCATCACTATCTCCAGCTACATCACTTCTCCACAACACACA 64

Oy 507 gataaccccaaaatcacacaagcgctcgtatgcgaccagaanaaacgatcccatct 566
| | | | | | | | | | | | | |
Db 65 CACAACCTTCTCCAGTCATCATCATTTCCAACAAACAACAACAACACTTTCTCACGTA 124

Oy 567 ggglaacatccacgcctcttaataacatcatgttccaactggaaggtgttcgtgacactca 626
 Db 125 CATCACTTCTCCAGCATCATCACTTCTCCAGCTACATCACTTCTCCAGCTCAACAA 184

627 ccgcctacatctcygaltcaaaactacttcaatctgttcgcagaaagaactgaaacgaagaat 686
Oy
185 cttctccagctcatcatcaattctcccaacagacaaacaaacaaactttctccaaacaaacca 244
Oy

[illegible]

Oy 747 cctgcagtagacaacaccgtactacatcgtgaattcgtaacgataccagaacaatatcgttga 806
 ||| ||| |
305 CATCAATATTTCCTAACACGCACAAACAAGTTTCCGCGTGTAATAAAGAATCATCCTGGCCCAC 364

QY	807	cotcaca	815
	1	11111	
365	CTACACCA	373	

RESULT 2
 CUSCOOL.CC

LOCUS	CNS00L00	1101 bp	DNA	linear	GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TERT3 end of BAC: BACR32D23 of RCE1-98 library from Drosophila melanogaster (fruit				

ACCESSION	AL066867	GI:4958669
VERSION	AL066867.1	
KEYWORDS	GSS.	

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE	1	mascomorphia, ephialoidea, diosiphallidae, diosiphallia.
AUTHORS	1 (bases to 1101)	
TITLE	Genoscope.	
Direct Submission	Direct Submission	

COMMENT
Determination of this RFLP and sequence was carried out as part of a project supported by the Centre National de Séquençage (E-mail: segreff@genoscope.cns.fr)
- Web: www.genoscope.cns.fr
BP 191 91106 Evry cedex - FRANCE

The BDGP is constructing a physical map of the *Drosophila melanogaster* genome, using these BACs. For further information please contact bdgp@berkeley.edu.

melanogaster BAC library was prepared by Kazuhiro Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,

Ecort digestion of *Drosophila* DNA provided by the BDP¹ from the isogenic strain y2; cn bw^{sp}, the same strain used for the BDP's P1 and EST libraries. A more detailed description of the library

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.utoronto.edu/drosophila_bac.htm.

Location/Qualifiers

1. 1101

source

Drosophila melanogaster

BASE COUNT	ORIGIN
155 a	166 c
7 g	284 t
489 others	

Query Match	4.1%;	Score 54.6;	DB 12;	Length 1101;
Best Local Similarity	17.6%;	Pred. No. 0.004;		
Matches 100;	Conservative 219;	Mismatches 245;	Indels 3;	Gaps 1

DQ 25 tctccctcactgataatcataagaacatcatcaataactgcattcgtaaccgtgcctac 84
TCVMACTATGCMHAAACMMHNAACMTAKTMMKCTMYATCCTHTYCHMHCMTGHMTAM 557

Qy 85 gaaccacatcactgcagtcgacctgtcttcgcctaagcttcccaaatcaacatcggtttctaa 144
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
b 558 cmttcatcmtcmtnnaaycmtscmctcttcmmcannttmamcamtmaanaacttttnc 617
:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy 145 gtaacttcgacgcagatcgaacaagaatcatccagctgttcaatctggaaattccaaa 204
+ :
bb 618 mmttttatmccncaahntmahnttmamhaahnnvttrrrmmmmmayahcacwvtttcaa 677
+ :

QY 205 atcgaagttatcctgaagaatgctatcgtatatacaactatgtaagaaaaccttcacac 264
::: || : : : | : : : || : : : ||
678 MIMCAACMHHMMMMHMTATMCCMMMAAAMAMMTTWTW---AMACMRAMMMCC 734

cy tctctcgcgattccgaataactccaatccatctctcgaacaatgaatacac 324
+ :
735 cscannmamaaaahmamcmamcmcmcmcmcmcmcmcmcmcmcmcmcmcm 794
+ :

[illegible][illegible]

445 atgatacaatctctgactatcataatcgcgtgagatctctgcgtttaccatcaaccacaatcgt 504
 :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
 505 atgatacaatctctgactatcataatcgcgtgagatctctgcgtttaccatcaaccacaatcgt 564
 :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

Qy 505 ctgaataactccaanaatctacatacgaagcgctgtgatlcatgcagaacgaatcttccaat 564
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : | : : : : :
+-----+

QY 565 ctgggtaacatccacgcgttctaataac 591
: ||:: :| |:: :||:
||:: :| |:: :||:

```
D0      100% HHHHHHPPPHHC I WCHNII PPRRHHIIR I    100%
```

RESULT 3

LOCUS	DEFINITION
BM277996	537 bp mRNA linear EST 20-DEC-2007
As tgz_51f02	SKPL Ascaris suum adult male testis germinal zone from tgz_51f02 5' mRNA sequence
Alao Scott	Ascaris suum cDNA clone As tgz_51f02 5' mRNA sequence

ACCESSION BM277996
VERSION BM277996.1 GI:17971254
KEYWORDS EST.
SOURCE

ORGANISM *Ascaris suum*
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
; Ascarididae; Ascaris.

AUTHORS
Blaxter, M. L., Parkinson, J., Whittton, C., Daub, J., Gulliano, D., Hall,
'N', Quayle, M. and Barrell, B.

TITLE Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL Unpublished (2000)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk

The library was prepared by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore. Sequencing was performed by the Pathogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell).

PCR Primers

FORWARD: T3
BACKWARD: T7PL
Plate: 51 row: F column: 02
Seq primer: SKPL
High quality sequence stop: 478.

FEATURES
source Location/Qualifiers

1..537
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone_lib="As_tgz_51F02"
/clone_lib="Ascaris suum adult male testis germinal zone from Alan Scott"
/sex="Male"
/dev_stage="Adult"

/note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI; Library was made from dissected testis germinal zone from adult male Ascaris suum collected from abattoirs. Constructed by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore, MD."

BASE COUNT 193 a 149 c 72 g 123 t
ORIGIN

Query Match 4.0%; Score 53.6; DB 10; Length 537;
Best Local Similarity 48.6%; Pred. No. 0.0055;
Matches 180; Conservative 0; Mismatches 184; Indels 6; Gaps 1;

QY 368 tgaactcgtggaatcctctggaactctgaggaactcaggaatcaacagcgtgtg 427
DB 49 tGTACACCAATGTCACAGAAACCACTGACAACTTCACACAGCAACTTTT 108
QY 428 tatcaatctctagatgatcaacatctctgactaactaatcgctgtgactctg 487
DB 109 GCAACTTCACAGTAATTCACACAGCAAGCCATTCACAGCAACTTCACATA 168
QY 488 ccatcaccaacatcgtctggaatcctcaaatctacatacagcgccgtctgacgacc 547
DB 169 ACAACTTAGCACTCTGAAACAACTTCAATTTTTCGCAACTCTGAGACACG 228
QY 548 agaacgcgactcccaatcgtggaatcagcgtcttaatacatalcgttcaaaagg 607
DB 229 GGAG-----CTACAACTTGACACACCTCAACACACTCAATACACCGAATCTCTA 282
QY 608 acggtgtcgtggaactcagcgtctacatctggtatcaataactcaatctgttcgacaag 667
DB 283 ACAACAACTTCAGCAATTCACAACTTCACAACTTCAGCAAGCAACTTCACATA 342
QY 668 aactgaagaaaagaatcaaaagcctgttagaacaacagtcgaattctgtatcctga 727
DB 343 ACAGTAATTCACAACTTCACAGCAACAACTTCAGCAACTTCACAACTTCACATA 402
QY 728 aagactctg 737
DB 403 ACAACAGCAG 412

RESULT 4
BM278689

LOCUS BM278689 551 bp mRNA linear EST 20-DEC-2001
DEFINITION As_tgz_66A05.SKPL Ascaris suum adult male testis germinal zone from Alan Scott
ACCESSION BM278689
VERSION BM278689.1 GI:17971947
KEYWORDS EST.
SOURCE pig roundworm.
ORGANISM Ascaris suum
; Ascarididae; Ascaris.

REFERENCE 1 (bases 1 to 551)

Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guillano,D., Hall

,N., Quayle,M. and Barrell,B.

Edinburgh University/Sanger Centre Nematode EST Project

Unpublished (2000)

CONTACT: Blaxter ML

Institute of Cell, Animal and Population Biology

University of Edinburgh

Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9

3JT, UK.

Tel: +44 131 650 6760

Fax: +44 131 670 5450

Email: mark.blaxter@ed.ac.uk

The library was prepared by Michelle Lizotte-Waniewski for Alan

Scott, Johns Hopkins University Medical School, Baltimore.

Sequencing was performed by the Pathogen Sequencing Unit, Sanger

Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell).

PCR Primers

FORWARD: T3

BACKWARD: T7PL

Plate: 66 row: A column: 05

Seq primer: SKPL

High quality sequence stop: 510.

FEATURES

source Location/Qualifiers

1..551
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone_lib="As_tgz_66A05"
/clone_lib="Ascaris suum adult male testis germinal zone from Alan Scott"
/sex="Male"
/dev_stage="Adult"

/note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI; Library was made from dissected testis germinal zone from adult male Ascaris suum collected from abattoirs. Constructed by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore, MD."

BASE COUNT 208 a 160 c 64 g 119 t
ORIGIN

Query Match 4.0%; Score 53.6; DB 10; Length 551;
Best Local Similarity 49.6%; Pred. No. 0.0055;
Matches 137; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY 437 acctcgaatgataacatctctgactacataatcgctggatctctgtacatcacca 496
DB 102 ACAGTAATTCACAACTTCACAACTTCACAACTTCACAACTTCACAACTTCACAA 161
QY 497 acaatcgtctgataaactccaaatctacatacagcgccgtctgactgacgcaaaacgga 556
DB 162 GCAACTCTGGAACAACTTCATTAACAACTTCAGCAACTCTGAGCAACCTCAGCTACA 221
QY 557 tctcaatctgggtgaacatcagcgtcttaataacatcgttcaaaactgacggtgtgc 616
DB 222 ACTTGAAACAACTTCACAACTTCACAACTTCACAACTTCACAACTTCACAACTTCAC 281
QY 617 gtgaacatcagcgtctacatctgataatctacatctgttcgacaagaagcgaag 676
DB 282 GTCTCAACATTAACCTTCACAACTTCAGCAACGCAATTTTCACAACTTCACAACTTCAC 341
QY 677 aaaaagaatcaagacctgttagacaacagctcca 712

Db	342	ACAGTAACCTGACGAACCTTCAACGACCACAACTTTCA	377
RESULT	5		
AU060224			
LOCUS	AU060224	394 bp	mRNA linear EST 20-MAY-1999
DEFINITION	AU060224 Dictyostellium discoideum SL (H.Urushihara) Dictyostellium		
ACCESSION	discoidem cDNA clone SLA610, mRNA sequence.		
VERSION	AU060224		
SOURCE	AU060224.1 GI:4881328		
ORGANISM	EST.		
KEYWORDS	Dictyostellium discoideum.		
REFERENCE	Eukaryota; Mycetozoa; Dictyostelidae; Dictyostellium.		
AUTHORS	Morito,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B.N., Pl.M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.		
TITLE	Developmental cDNA in Dictyostellium discoideum		
JOURNAL	Unpublished (1998)		
COMMENT	Contact: Hideko Urushihara Institute of Biological Sciences University of Tsukuba 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan Email: d402hu@tsakura.cc.tsukuba.ac.jp PROJECT - Dictyostellium discoideum cDNA project in Japan. Location/Qualifiers 1..394 /organism="Dictyostellium discoideum" /strain="AX4" /db_xref="taxon:44689" /clone="SLA610" /clone_id="Dictyostellium discoideum SL (H.Urushihara)" /dev_stage="slug"		
BASE COUNT	223 a 134 c 8 g 28 t 1 others		
ORIGIN			
Query Match	4.0%; Score 53.2; DB 9; Length 394;		
Best Local Similarity	49.3%; Pred. No. 0.0061;		
Matches 139; Conservative	0; Mismatches 143; Indels 0; Gaps 0;		
OY	432 caaatcctcgtagatcaacatcttcgactacatacgtctgatcttgcattcat 491		
Db	111 CAACACGCACCAACAACACCAACCAACAACAACAACAACAACAACAACA 170		
OY	492 caccacaacatcgcttgaaataacctccaataatctacatacaaggccgctcgatcgaccagaa 551		
Db	171 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 230		
OY	552 accgagctccaatctcggttaatatcccaagctctctaatacatcatgttcaaatcgtgaegg 611		
Db	231 TATCATTCACACATCAGCAACAATATATCATCATATATCAACAACAATFACCATCCATCACAA 290		
OY	612 ttgctgtgacctcacctccgttatcatcttgatcataacttcaatctcgttcggaccagaact 671		
Db	291 CAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 350		
OY	672 gaacgaaaaagaatatcaaagaccttgaagacaaccaagtcca 713		
Db	351 CAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 392		
RESULT	6		
BIS94803			
LOCUS	BIS94803	447 bp	mRNA linear EST 07-SEP-2001
DEFINITION	As.tgz_44C11.SKPL Ascaris suum adult male testis germinal zone from		
ACCESSION	Alan Scott Ascaris suum cDNA clone As.tgz_44C11 5', mRNA sequence.		
VERSION	BIS94803		
KEYWORDS	BIS94803.1 GI:15498290		
SOURCE	EST.		
ORGANISM	pig roundworm.		
	Ascaris suum		

```

REFERENCE 1
AUTHORS Blaxter/M.L., Parkinson,J., Whitton,C., Daub,J., Gulliano,D., Hall
          ,N., Quayle,M. and Barrell,B.
TITLE Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL Unpublished (2000)
COMMENT Contact: Blaxter ML
          Institute of Cell, Animal and Population Biology
          University of Edinburgh
          Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
          3J7, UK.
          Tel: +44 131 650 6760
          Fax: +44 131 670 5450
          Email: mark.blaxter@ed.ac.uk
          The library was prepared by Michelle Lizotte-Waniewski for Alan
          Scott, Johns Hopkins University Medical School, Baltimore.
          Sequencing was performed by the Pathogen Sequencing Unit, Sanger
          Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
          sequence contained a Polya tail (trimmed)
          PCR Primers
          FORWARD: T3
          BACKWARD: T7/PL
          Plate: 44 row: C column: 11
          Seq primer: SKPL
          High quality sequence stop: 447.
          Location/Qualifiers
          1..447
FEATURES
source
```

		/note=Vector: lambda Zap II; Site.1: EcoRI; site.2: XhoI;		
		Library was made from dissected testis germinal zone from		
		adult male Ascaris suum collected from abbattoirs.		
		Constructed by Michelle Lizotte-Waniewski for Alan Scott,		
		Johns Hopkins University Medical School, Baltimore, MD."		
BASE COUNT	173 a	127 c	49 g	98 t
ORIGIN				
Query Match	3.9%	Score 52.6;	DB 10;	Length 447;
Best Local Similarity	49.8%;	Pred. No. 0.0092;		
Matches 133;	Conservative 0;	Mismatches 134;	Indels 0;	Gaps 0;
Db	3	TGACACAAGCAATTTTCACACAGCAACTGCAGCACTTAATTAAACACTTGTGACAACTCTC	62	
Qy	446	tgaataacatctcggactacaatcgctgcatcttcgtaccataccaacaatgctc	505	
Db	506	tgaataactccaanaattcacaaaggcgctcgtatcgacagaagaacggaatcccaatc	565	
Db	63	TGAACAACCTTGCAATTAACAACTCCAGCAACTCTCTGAGCAACCTCGATYAAACCTTGACAA	122	
Qy	566	tgggttaaatcatacagcttctaataacatcatcgttcaaacatcgtagtgctgtagaacatc	625	
Db	123	ACCTTCACAACCTTCACAACAACCTCAACTAATCTTAATAAACATCTCAGCATCTCAACA	182	
Qy	626	acgcgtacatcttgatacaaatacttcaatctcgttcgcagaagaagtgaacgaaaaagaaa	685	
Db	183	ATAAACCCTCAACAACTTGACAAACAGCAAGCAATTTCAACAGCAGATTAATTCAACAACAGTAGTACT	242	
Qy	686	tcaaagacctgtaegacaaccaagtcca	712	
Db	243	TCAGCAACTTCACAGCAACGAAACACTTCA	269	
RESULT	7			
LOCUS	BM278181	597 bp	mRNA	linear EST 20-DEC-2001
DEFINITION	As_tg_54C05.SKPL Ascaris suum adult male testis germinal zone from			

ACCESSION	ALAN SCOTT Ascaris suum cDNA clone As_tgz_54C05 5', mRNA sequence.
VERSION	BM278181
KEYWORDS	BM278181.1 GI:17971439
SOURCE	EST.
ORGANISM	pig roundworm. Ascaris suum
REFERENCE	Eukaryote; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea ; Ascaridae; Ascaris. 1 (bases 1 to 597)
AUTHORS	Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Gulliano,D., Hall N., Quayle,M. and Barrrell,B.
TITLE	Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL	Unpublished (2000)
COMMENT	Contact: Blaxter ML

Institute of Cell, Animal and Population Biology
 University of Edinburgh
 Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
 3JT, UK.
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark_baxter@ed.ac.uk
 The library was prepared by Michelle Lizotte-Maniewski for Alan
 Scott, Johns Hopkins University Medical School, Baltimore.
 Sequencing was performed by the Pathogen Sequencing Unit, Sanger
 Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell).
 PCR Primers
 FORWARD: T3
 BACKWARD: T7PL
 Plate: 54 row: C column: 05
 Seq primer: SKPL
 High quality sequence stop: 502.

```

FEATURES
source
  location/Qualifiers
    1..597
      /organism="Ascaris suum"
      /db_xref="taxon:6253"
      /clone="As_t92_54C05"
      /clone_11b="Ascaris suum adult male testis germinal zone
        from Alan Scott"
      /sex="Male"
      /dev_stage="Adult"
      /note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI.
        Library was made from dissected testis germinal zone from
        adult male Ascaris suum collected from abattoirs.
        Constructed by Michelle Licotte-Waniewski for Alan Scott,
        Johns Hopkins University Medical School, Baltimore, MD."
BASE COUNT
222 a 167 c 76 g 132 t

```

Query Match	3.9%	Score 52;	DB 10;	Length 597;
Best Local Similarity	49.3%	Pred. NO. 0.015;		
Matches 136; Conservative	0;	Mismatches 140;	Indels 0;	Gaps 0;

[illegible]

RESULT	8
BM278805	
LOCUS	500 bp mRNA linear EST 20-DEC-2001
DEFINITION	As_t9_67F10.SKPL Ascaris suum adult male testis germinal zone from Alan Scott Ascaris suum cDNA clone AS_t9_67F10 5', mRNA sequence.
ACCESSION	BM278805
VERSION	BM278805.1 GI:17972063
KEYWORDS	EST.
SOURCE	pig roundworm,
ORGANISM	Ascaris suum

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JF, UK.
Tel: +44 131 650 5760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared by Michelle Lizotte-Waniewski for Alan
Scott, Johns Hopkins University Medical School, Baltimore.
Sequencing was performed by the Pathogen Sequencing Unit, Sanger
Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
sequence contained a Polya tail (trimmed)

```

PCR primers
FORWARD: T3
BACKWARD: T7PL
Plate: 67 row: F column: 10
Seq primer: SKPL
High quality sequence stop: 500.
location/Qualifiers
1. .500
source

```

```

/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone-"As taz.67F10"
/clone_lib="Ascaris suum adult male testis germinal zone
from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/notes="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI;
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Licotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."

```

	Query Match	3.9%	Score 51.8;	DB 10;	Length 500;	
	Best Local Similarity	50.8%;	Pred. No. 0.015;			
	Matches 153;	Conservative 0;	Mismatches 142;	Indels 6;	Gaps 1:	
Qy	437	actctcagaatgaacactctctgacatacatcaatcgctgatcttcgtttaccacaccca	496			
		" "				
Db	70	ACAGTAACCTTGAACCAACGACATTTTCACACACCACCTCAGAACCTCATATACAACCTTA	129			
Qy	497	acaatcgcgtgaataactcccaaaatlctacccaagcgctctgatacgaccaagaaccca	556			
		" "				
Db	130	GCAACTCTCTGAAACAACCTTCATTAACCAACCTCAGAACCTCTGTGGCAACCTCAG----	184			
Qy	557	tctccaattctggtaaacctcaacgccttctaataacatcatgttcaaactggagtgttc	616			
		" "				
Db	185	-CTAACAATTGAAACAACCTCAACAACCTCATTAACCAACCTCAACATATCTCAACAACATTC	243			
Qy	617	gtgaacctcacgcgtacatatcgtalccaatlaacttaactcgttctgcagaaagaactgaacg	676			
		" "				

[illegible]

Query Match	3.8%	Score 50.6	DB 10	Length 511
Best Local Similarity	46.9%	Pred. No. 0.031		
Matches 158	Conservative 0	Mismatches 179	Indels 0	Gaps 0
QY	432	caaatcactcctagatgatacaacatcttcgtactacatacgaatcgctgtagatcttgaccat	491	
DB	169	CTACTACTACAACTACGACACCAACCAACTACACAACTACACAACTCTACACTACACCAACAA	228	
QY	492	caaccaaatcgtctcgtataatccaaatctcaatcaaaagggccggtctgcagccagaa	551	
DB	229	CAACGACACAAACCAACCAACTACACCAACAAACGACGACAAACCAACCAACTACAC	288	
QY	552	accgattcccaatctcgtgtaacatccacgcctctcaataacatggtccaactggagcg	611	
DB	269	CAACACACACGACACACACACACGCACTACACGACACATCTACGACACCAACCACTACGA	348	
QY	612	ttgtcgtgacactcaccgcgtcatcattcgtgataaatacttcaatctgttcgcaagaact	671	
DB	349	CTACCCACAACTACACTACACCAACTACACCAACGACAAACGACCAACCAACTACAA	408	
QY	672	gaacgaaaaagaatccaagacctgtacgacacaccagtcgaattctgtatcccgaaaga	731	
DB	409	CAACACACACGACGACCAACCAACCAACTACACGACGACGACGACGACGACGACGACG	468	
QY	732	ctctcggggtgactaccctgcagtagcagaaacccgtac	768	
DB	469	CTACGACACACATCTACGACGACATCTACACACGACGACGACGACGACGACGACGACG	505	
RESULT 10				
LOCUS	BM278657			
DEFINITION	As.tgz.65E06.SKPL Ascaris suum adult male testis germinal zone from Alan Scott Ascaris suum cDNA clone As.tgz_65E06.5', mRNA sequence.			
ACCESSION	BM278657			
VERSION	BM278657.1			
KEYWORDS	EST.			
SOURCE	pig roundworm.			
ORGANISM	Ascaris suum			
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea			
AUTHORS	1 (bases 1 to 569)			
TITLE	Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Gulliano,D., Hall			
JOURNAL	,N., Quayle,M. and Barrell,B.			
COMMENT	Edinburgh University/Sanger Centre Nematode EST Project , Unpublished (2000) Contact: Blaxter ML Institute of Cell, Animal and Population Biology University of Edinburgh Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK. Tel: +44 131 650 6760 Fax: +44 131 670 5450 Email: mark.blaxter@ed.ac.uk The library was prepared by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore. Sequencing was performed by the Pathogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The sequence contained a Polya tail (trimmed) PCR primers FORWARD: T3 BACKWARD: T7/PL Plate: 65 row: E column: 06 Seq primer: SKPL High quality sequence stop: 518. Location/Qualifiers 1..569 /organism="Ascaris suum" /db_xref="taxon:6253" /clone="As.tgz_65E06" /clone_lib="Ascaris suum adult male testis germinal zone from Alan Scott"			

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 21:29:41 : Search time 7254.67 Seconds
(without alignments)
3816.274 Million cell updates/sec

Title: US-09-611-419A-3

Perfect score: 1323

Sequence: 1 gaattcgaaacgatgtctac.....aacgtccgctgaagaattc 1323

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenEmbl:
1: gb-ba:*
2: gb-htg:*
3: gb-in:*
4: gb-om:*
5: gb-ov:*
6: gb-pat:*
7: gb-ph:*
8: gb-pl:*
9: gb-pr:*
10: gb-ro:*
11: gb-sts:*
12: gb-sy:*
13: gb-un:*
14: gb-vl:*
15: em-ba:*
16: em-fun:*
17: em-hum:*
18: em-in:*
19: em-mu:*
20: em-om:*
21: em-or:*
22: em-ov:*
23: em-pat:*
24: em-ph:*
25: em-pl:*
26: em-ro:*
27: em-sts:*
28: em-un:*
29: em-vl:*
30: em-htg-hum:*
31: em-htg-inv:*
32: em-htg-other:*
33: em-htg-inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
------------	-------------	-------	--------	-------	-------------

1	1304.4	98.6	1330	6	AR000029
2	1304.4	98.6	1330	6	AR169140
3	1304.4	98.6	1330	6	AX036243
4	1304.4	98.6	1338	12	XXU22962
5	1304.4	98.6	1402	6	AR000030
6	1304.4	98.6	1402	6	AR169141
7	1304.4	98.6	1402	6	AX036246
8	727.6	55.0	4292	1	CBNT02
9	727.6	55.0	4835	6	CLONEUR
10	726.4	54.9	3891	6	AR000031
11	726.4	54.9	3891	6	AR169142
12	726.4	54.9	3891	6	AX036248
13	662	50.0	4067	1	CBNT02A
14	423.2	32.0	1239	12	AF251281
15	421.6	31.9	1313	6	A58946
16	228	17.2	4073	1	CBNT02F
17	227	17.2	1293	6	A58945
18	227	17.2	4199	1	CLONEUR
19	227	17.2	4209	1	CBNT02G
20	195.8	14.8	3835	1	AB037704
21	195.8	14.8	3835	1	AB037705
22	195.8	14.8	3835	1	AB037706
23	195.8	14.8	3835	1	AB037707
24	195.8	14.8	3835	1	AB037708
25	195.8	14.8	3835	1	AB037709
26	195.8	14.8	3835	1	AB037710
27	195.8	14.8	3835	1	AB037711
28	195.8	14.8	3835	1	AB037712
29	195.8	14.8	3835	1	AB037713
30	195.8	14.8	3835	1	AB037714
31	195.8	14.8	4030	1	CBNT02E
32	194.2	14.7	4017	6	AX088262
33	194.2	14.7	4017	6	AX088264
34	191	14.4	3849	1	CBNT02F
35	180.8	13.7	4234	1	CLONEUR
36	180.8	13.7	9325	1	CBNT02F
37	180.8	13.7	9325	1	CBNT02F
38	166.4	12.6	1359	6	I28431
39	166.4	12.6	3712	6	A49987
40	166.4	12.6	3712	6	A49987
41	166.4	12.6	3754	6	A37074
42	166.4	12.6	3754	6	A42478
43	166.4	12.6	3769	6	A37075
44	166.4	12.6	3769	6	A49988
45	166.4	12.6	4366	6	A42484
			4378	6	A42481

ALIGNMENTS

RESULT 1
LOCUS AR000029 1330 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 22 from patent US 5736139.
ACCESSION AR000029
VERSION AR000029.1 GI:3962560
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1330)
AUTHORS Kink,J.A., Thalley,B.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.
TITLE Treatment of Clostridium difficile induced disease
JOURNAL Patent: US 5736139-A 22 07-APR-1998;
FEATURES
source
1. 1330
/organism="unknown"
BASE COUNT 400 a 339 c 246 g 345 t
ORIGIN

Query Match 98.6%; Score 1304.4; DB 6; Length 1330;
Best Local Similarity 99.9%; Pred. No. 0;

SEQID 3

Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 12 gatgtcacttcactgaatacatcaagaacatcatcaactcctcctgaacccg 71
   |||
Db 12 GCTGTACTTCACTGATATATCATCAAGAACATCATCAATCACTCCATCTCGAAGCTGCG 71

QY 72 ctacgaatccaatcaactgaatcgactctcgtctacgcttccaaatcaacatcg 131
   |||
Db 72 CTACGAATCCAAATCACCTGATCGACTCTCCGCTACGCTTCCAAATCAACATCGGTTTC 131

QY 132 taaagtaacttcgatccgctgaagaagaatgaatcgaatcgaatcgaatcga 191
   |||
Db 132 TAAAGTTAACTTGCATCGATCGACAAAGATCGATCGATCGATCGATCGATCGATCGAT 191

QY 192 caaatcgaaatgaatcgaagaatgctatcgatatacaactctatgaacgaactctc 251
   |||
Db 192 CAAATCGAAGTTATCTGTGAAGATGCTATCGATATCACTATGATGACGAATCTTCTC 251

QY 252 caactcctctgagtcgctatccgaaatacttcaactcctcctcgaacatgata 311
   |||
Db 252 CACCTCTCTGTGATCGCTATCCGAAATCTTCAACTCCATCTCTGGAACATGATA 311

QY 312 caccatcaactgaatcgatgaagaacatcttggttgaagaatcctgaactgaag 371
   |||
Db 312 CACCATATCAACTGATGGAAGAACATCTGTTGGAAGATCTGTGAAGTACGATG 371

QY 372 aatcatctgagctcgaagacactcgaagaatcaaacagcgctgtgtatccaactc 431
   |||
Db 372 AATCATCTGAGCTCTGAGACACTCGAAGAACATCAACAGCGTGTGTATTCAAATCTC 431

QY 432 tcagatgaatcaacatctcgtactacatcaatcgtcgtatcttgatccaacacaa 491
   |||
Db 432 TCGATGATCAACATCTCTACTACATCAATCGCTGATCTCTGATCCATCAACAA 491

QY 492 tcgtctgaatcaactccaactctacatcaacgagcgctcgtatcgacgaagaac 551
   |||
Db 492 TCGTCTGAATCACTCCAAATCTATCATCAACGCGCTCTGATCGACGAAGACGATCTC 551

QY 552 caatctgagtgaaatcaatccagccttctaaatacatcatgttcaactgagcggt 611
   |||
Db 552 CAATCTGGTAACTCATCCACTCTTATATCATCATGTTCAAACTGAGCGTGTCTGGA 611

QY 612 cactcaccgctatcatcgtggaatcaactcctcgttctggaagaactggaacgaaa 671
   |||
Db 612 CACTCACCGGTACTCTGATCAATCAATCTTCAATCTTGTGACAAACAACTGGAAGAAA 671

QY 672 agaatcaagaagcctgacgaacaaacagtcgaatctcgtatcctcgaagaactc 731
   |||
Db 672 AGAATCAAGAGCTGTAGACAAACAGTCAATCTGATCTCTGGAAGACTCTTGCGG 731

QY 732 tgactacatcgaatgaagaacacgcttactatcgtatcgtatcgtatcgtatc 791
   |||
Db 732 TGACTACTCTGATGACAAACAGTCAATCTGATCTGATCTGATCTGATCTGATCTGAT 791

QY 792 cgttgaagtgaaatgaatgaatcgtatcgtatcgtatcgtatcgtatcgtatc 851
   |||
Db 792 CGTTGAGGTCAACATATGATATCCGCGGTATCATGTACCTGGAAGCTCGCGTGTTC 851

QY 852 tctatcgaacccaactcactcgaactctccctgtacgctgtgtacgaatcgaatc 911
   |||
Db 852 TGTATGATCAACATATGATATCCGCGGTATCATGTACCTGGAAGCTCGCGTGTTC 911

QY 912 caagaatcagcgctgtgtaagaagaacatctcgtcgaacaaatgactgtatatac 971
   |||
Db 912 CAAGAATACGCGCTGTGTAAACAGGACAAATATGTTGCGAACAAATGATGCTGATACAT 971

QY 972 caatgtgtgtgtgaagaagaagaatcgtcgtcgtacgaatcgtcgtcgtcgtcgt 1031
   |||
Db 972 CAATGTGTGTGTGAAGAACAAAGATACCGCTGTGTGACCAATGCTTCTCAAGCTGTGTGT 1031

QY 1032 agaaaagatctgtcgtcgtgaacacccgagcgtgtgtaactcgtcgtcgtcgtcgt 1091
   |||
Db 1032 AGAAAAGATCTGTCTCTCTGGAATCCGCGAGCTGTGTGATATGCTGCTCAAGTATGTTGT 1091

```

```

QY 1092 aatgaatccaagaacgaccaggtatcactaaacaaatgcaaatgactcgaagca 1151
   |||
Db 1092 AATGAATCCAAGAACACAGCGGTATCACTAACAATGCAAAATGATATGCGAGCA 1151

QY 1152 caatgtaacgatacggcttccatcggcttccacagttcaacaaatcgtcgaactcgt 1211
   |||
Db 1152 CAATGTGATGATATCGTTTCAATCGGTTCACCACTTCAACAAATATCGCTAAACGTGT 1211

QY 1212 tgcctcgaactgtgataatcgtcgaatcgaactcgtcgtcgtcgtcgtcgtcgt 1271
   |||
Db 1212 TGCCTCCAACTGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1271

QY 1272 ggaatcaccggtgtgatcaggtgtggtgtgaacgtcgtcgtcgtcgtcgtcgt 1317
   |||
Db 1272 GGAATCATCCCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317

RESULT 2
AR169140 1330 bp DNA linear PAT 17-DEC-2001
LOCUS AR169140
DEFINITION Sequence 22 from patent US 6290960.
ACCESSION AR169140
VERSION AR169140.1 GI:17906909
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 1330)
AUTHORS Kink,J.A., Thaller,B.S. and Stafford,D.C.
TITLE Vaccine and antitoxin for the treatment of C. difficile disease
JOURNAL Patent: US 6290960-A 22 18-SEP-2001;
FEATURES
source 1..1330
BAPF COUNT 400 a 339 c 246 g 345 t
ORIGIN

Query Match 98.6%; Score 1304.4; DB 6; Length 1330;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 gatgtcacttcactgaatacatcaagaacatcatcaactcctcctgaacccg 71
   |||
Db 12 GCTGTACTTCACTGATATATCATCAAGAACATCATCAATCACTCCATCTCGAAGCTGCG 71

QY 72 ctacgaatccaatcaactgaatcgactctcgtctacgcttccaaatcaacatcg 131
   |||
Db 72 CTACGAATCCAAATCACCTGATCGACTCTCCGCTACGCTTCCAAATCAACATCGGTTTC 131

QY 132 taaagtaacttcgatccgctgaagaagaatgaatcgaatcgaatcgaatcga 191
   |||
Db 132 TAAAGTTAACTTGCATCGATCGACAAAGATCGATCGATCGATCGATCGATCGATCGAT 191

QY 192 caaatcgaaatgaatcgaagaatgctatcgatatacaactctatgaacgaactctc 251
   |||
Db 192 CAAATCGAAGTTATCTGTGAAGATGCTATCGATATCACTATGATGACGAATCTTCTC 251

QY 252 caactcctctgagtcgctatccgaaatacttcaactcctcctcgaacatgata 311
   |||
Db 252 CACCTCTCTGTGATCGCTATCCGAAATCTTCAACTCCATCTCTGGAACATGATA 311

QY 312 caccatcaactgaatcgatgaagaacatcttggttgaagaatcctgaactgaag 371
   |||
Db 312 CACCATATCAACTGATGGAAGAACATCTGTTGGAAGATCTGTGAAGTACGATG 371

QY 372 aatcatctgagctcgaagacactcgaagaatcaaacagcggtgtgtatccaactc 431
   |||
Db 372 AATCATCTGAGCTCTGAGACACTCGAAGAACATCAACAGCGTGTGTATTCAAATCTC 431

QY 432 tcagatgaatcaacatctcgtactacatcaatcgtcgtatcttgatccaacacaa 491
   |||

```

D	432	TCAGATGATCAACATCTCTGACTACATCAATCCGCTGATCTGTTACATCAACAA	491
Q	492	tcgtctgaataactccaaatactacatcaacggcgctctgaatgacgaacagatctc	551
D	492	TCGCTGAATAAATCTCAAAATCTACATCAACGGCGCTGATGACACGAACCGATCTC	551
Q	552	caatctggtaacatccacgcctctcaataacatctgtaaacatgagcgtctgctga	611
D	552	CAATCTGGTAACATCCACGCTTCTAATACATCATCTTCAAACTGACGCTTGTCTGA	611
Q	612	caactacgcgtacatctgatacaatactcaatctgtctgcacaagaactgacga	671
D	612	CACCTACCGCTACATCTGGATTCAAATCTCTTGTGACAAAGATGAAACGAAAA	671
Q	672	agaatacaagactgtgacgaacacacgctcaatcttcgfatctctgaaactctggg	731
D	672	AGAAATCAAAAGACCTGTGACACAAACAGTCCAAATCTGATATCTGAAAGATCTGGG	731
Q	732	tgactactgcaatgacgaacacacgctcaatctgtaacatctgtaacacga	791
D	732	TGACTACTGCACTGACGACAAACCGTACTACATGCTGAATCTGTAGATCCGAACAATA	791
Q	792	cgttgacgttaacaatgataagttacggtttacatgtaacgtgaagtcggtgtc	851
D	792	CGTTGACGTCAACATGATAGGTATCCGCGTTACATGATACCTGAAAGGTCCGCGTTC	851
Q	852	tgatatactacaacatctacacgtgaactcttcctgtacacgtgtaacgaatctcat	911
D	852	TGTTATACTACCAACATCTCACTGAACTCTCCCTGTACCGTGATGCAAAATTCATCAT	911
Q	912	caagaataacgctctgtaacgaacgaatctgctgcaacatgatactgatacat	971
D	912	CAAGAATACGCGTCTGTGTAACAAAGCAATATCGTTGCAACATGATGATATCAT	971
Q	972	caatgtgtgataagaacgaacgaatcgtctgctacatgctctcaagctgtgt	1031
D	972	CAATGTGTGATTAAGAAACAAAGAAATACCGTCTGCTACCAATGCTCTCAGCGTGTGT	1031
Q	1032	agaagaatcttctctctctgtaaatcccggaacgttgtaactctgctgaagttgt	1091
D	1032	AGAAAGATCTTGTCTCTCTGGAATCCCGGAGCTGTGTAATCTGCTCAAGGATGTGT	1091
Q	1092	aatgaataccgaacgaacgaacggtatcactaaatgcaaaatgaaatctgcagaca	1151
D	1092	AATGAATCCAAAGAACGACGAGGTATCACTAACAATGCAAAATGATTCGACGACAA	1151
Q	1152	caatgtgaacgatacggttcatcggttccacagttcaacaatacgttaactggt	1211
D	1152	CAATGTGATGATTCGCTGTTCAATCGGTTCCACAGTTCACAAATATTCGCTAAACTGCT	1211
Q	1212	tgcttcaactgtgataatgtaacatcgatcgaaactctctcgcacgtctggtgtgt	1271
D	1212	TGCTTCAACTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1271
Q	1272	ggagttcatccggttgatgacggttggtggaacgctgcgctgtga	1317
D	1272	GGAAGTTCATCCCGGTGATGACGCTTGGGATGAACGTCCTGCTGAA	1317
RESULT	3		
LOCUS	AX036243	1330 bp	DNA
DEFINITION	Sequence 22 from Patent EP1041149.		Linear
ACCESSION	AX036243		
VERSION	AX036243.1	GI:11225862	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE	1 (bases 1 to 1330)		
AUTHORS	Kink,J.A., Flicha,J.R., Padhye,N.V., Thalley,B.S., Stafford,D.C. and Williams,J.A.		
TITLE	Vaccine and antitoxin for treatment and prevention of C. Difficile		

JOURNAL	Patent: EP 1041149-A 22 04-OCT-2000;	OPHIDIAN PHARM INC (US)	Location/Qualifiers
FEATURES	source	1..1330	/organism="synthetic construct"
CDS	1..1317	/db_xref="taxon:32630"	
	/note="Synthetic"		
	/codon_start=1		
	/transl_table=11		
	/protein_id="CAC16478.1"		
	/db_xref="GI:11225863"		
	/translation="MARLSTFEYIKNTINTSIILNRYESNHLIDLSRVASKINIGS KYNPDIDKNOIOLNLSSEKTEVILKNAIYVNSMVENSTSFRIKRYPSISLN EYTIINCMENNSGKRYSLNGEIIITLDTDTGRTKRYPKYSOMINISDIYINRAIYFT ITNNRLNNSKIITINGLIDQKFIISNGIHNASHNMFILKDGDRDRTHTWIKYELFD KELNEIKEDLYDNOSNSGILDFWGDYLYQDKPYMMNLXDPNRYVNVNNGIGYM YLKPGSVMTNINILNSLYRGTFEIIKKYVSKNDIVRNNDVYINVVKNKEFR LATNASQAVEKILSALIEIPDVGNLSQVVMKSKMDGITTCKKNLDDNNGNDIGFI GFHOENNIKKLVASNMWYNRQIRRSSTLTGCSNEFIIPVDGMEERPL"		
BASE COUNT	400 a	339 c	246 g 345 t
ORIGIN			
Query Match	98.6%; Score 1304.4; DB 6; Length 1330;		
Best Local Similarity	99.9%; Pred. No. 0;		
Matches 1305; Conservative	0; Mismatches 1; Indels 0; Gaps 0;		
Q	12	gatgtctaccttactgaatacacaagaacatcaatcctcatcctgaactgcg	71
D	12	GCTGTACTCTACGATGATACATCAAGAACATCATCAATCACTCCATCCTGAACCTGG	71
Q	72	ctagaatccaatcactgactgactgtctgctgactgactgactgactgactgactg	131
D	72	CTAGGAATCCAAATCACCCTGATTCACCTGTCTGCTACGCTTCCAAATCAACATCGGTT	131
Q	132	taagttacttgcattcgcgtgacgaagaatcagatccacgtcttcaatctgaatcttc	191
D	132	TAAAGTTAACTTGGATCCGATCCGATCAAGAAATCGATCCAGCTTTCAATCTGGAATCTTC	191
Q	192	caaatcgaagtatccctgaagaatgctatcgtatacaactctatgtaacaaactctc	251
D	192	CAAAATGAGATTAATCCTGAAGATGCTATGATCAACTATATGACGAAATCTTCTC	251
Q	252	caactcctctgatacgtatcccgaaatactcaactcctcctcgtgaacaaatga	311
D	252	CACCTCCTTCTGATCCGATCCCGAATATCTCAACTCCATCTCTCTGAACAAATGATA	311
Q	312	cacatcatcaactgacatggaagaatctgttggaagaatctctctgaactgactg	371
D	312	CACCATCATCAACTGCATGGAAGAAATCTGTTGGAAGATATCTCTGACATCAAGCTGA	371
Q	372	aatcatctgactctgcaagacacacgaagaatcaaacagcgtgtgtatcaatactc	431
D	372	AATCATCTGACATCTGAGGACACTCAGGAATCAACACGCTGTATTAATCAATATCTC	431
Q	432	tcagatgatacaaatctgactacataatgctgtgattcttggttaacatacaaca	491
D	432	TCAGATGATCAACATCTCTACTACATCAATCCCTGATCTTGTACCAATCAACAA	491
Q	492	tcgtctgaataactccaaatactacatcaacggcgctctgatacgaacgaacgctctc	551
D	492	TGCTTGAATAAATCTCAAAATCTACATCAACGGCGCTGATGACACGAACCGATCTC	551
Q	552	caatctggtaacatccacgcctctcaataacatctgtaaacatgagcgtctgctga	611
D	552	CAATCTGGTAACATCCACGCTTCTAATACATCATCTTCAAACTGACGCTTGTCTGA	611
Q	612	caactacgcgtacatctgatacaatactcaatctgtctgcacaagaactgacga	671
D	612	CACCTACCGCTACATCTGGATTCAAATCTCTTGTGACAAAGATGAAACGAAAA	671

QY 672 agaaatcaagaacctgtacgaacaaccagtcacattctgtatccttgaaagactctctggg 731
 Db 672 AGAATTCATAAGACCTGTACGACCAACCAATTCCTGATCCCTGAAAGACTTCTGGGG 731
 QY 732 tgactactcagctagcagaacacgctactactatctgaatctgtacgactcgaacaata 791
 Db 732 TGACTACTCGAGTACGACCAACCGCTACTACATGCTGATCTGACGATCCGAACAAATA 791
 QY 792 cgttgacgtcaacaatgtaggatccgcgtgtacatctgacttgaaaggtccgcgtgtgtc 851
 Db 792 CGTGTAGCTCAACATGTAGGTATCCGGGTGTACATGTACTGAAAGTCCGGGTGTGT 851
 QY 852 tcttatgactcacaacatctactcgtgaactctctccctgtacccgtgtgacaaatcatcat 911
 Db 852 TGTATGTACTACCAACATCTACTCTGAACTCTCCCTGTACCGGTGTACCAATTCATCAT 911
 QY 912 caagaataacgctcgtgttaacaagaacatacgtctgcgaacaatgatacgtgtatcat 971
 Db 912 CAGAAATACGCGTGTGTTAACAAGACAAATATGCTGTGCAACAATGATCGTATACAT 971
 QY 972 caatgtgttagtlaagaacaagaataccgctgtgtctacaaatgtctctcagcgctgtgt 1031
 Db 972 CAATGTGTAGTTAAGAACAAGAATACGCTGTGGCTACCAATGCTTCTCAGGCTGTGT 1031
 QY 1032 agaaaagatctgtctgtctcgtcgtgaatcccggaagctgtgatactgtctcagctagtgt 1091
 Db 1032 AGAAAAGATCTGTCTGTCTGTGGAATCCCGACGTTGGTATCTGTCTCAGGTAGTGT 1091
 QY 1092 aatgaataccaagaacgacgaggtgtatcactaacaataatgaataatcgtcgaagaa 1151
 Db 1092 AATGAATACCAAGAACGACGAGGTATCTACTACAAATGCAAAATGATCTGCCAGGACAA 1151
 QY 1152 caatgttaacgatalcggcttccatcgtgttccaccagttccaacataatcgttaactgt 1211
 Db 1152 CAATGTTAACGATATCGGTTTCATCGGTTTCACACAGTTCAACAATATCGCTAAACTGT 1211
 QY 1212 tgcctccacacgtgtacaaatcgtcgaatcgttccctcgtcgaactcgtggtgtctgt 1271
 Db 1212 TGCTTCCACAGTGTACATCTGTCAATGCAAGCTTCTCTGCACTCGTGGGTGCTCTTG 1271
 QY 1272 ggaattcaccccggttgcagcgtgttggtggtgaacgctcgcgtctga 1317
 Db 1272 GGAGTCAATCCCGGTGTGATGACGGTGTGGGTGGAACGTCCGCTGTGA 1317
 ESURT 4
 LOCUS XU22962 1338 bp DNA linear SYN 08-NOV-1995
 DEFINITION Synthetic botulinum neurotoxin serotype A Hc fragment (botA) gene,
 complete cds.
 ACCESSION U22962.1 GI:733428
 VERSION U22962.1
 KEYWORDS synthetic construct.
 SOURCE synthetic construct.
 ORGANISM artificial sequence.
 REFERENCE 1 (bases 1 to 1338)
 AUTHORS Clayton, M.A., Clayton, J.M., Brown, D.R. and Middlebrook, J.L.
 TITLE Protective vaccination with a recombinant fragment of Clostridium
 botulinum neurotoxin serotype A expressed from a synthetic gene in
 Escherichia coli
 JOURNAL Infect. Immun. 63 (7), 2738-2742 (1995)
 MEDLINE 95310035
 REFERENCE 2 (bases 1 to 1338)
 AUTHORS Brown, D.R.
 TITLE Direct Submission
 JOURNAL Submitted (17-MAR-1995) Douglas R. Brown, Toxinology, U. S. Army
 Medical Research Institute of Infectious Diseases, Bldg. 1425, Fort
 Detrick, Frederick, MD 21702-5011, USA
 FEATURES
 source 1..1338
 /organism="synthetic Construct"
 /db_xref="taxon:32630"

/lab_host="Escherichia coli"
 /note="based on Clostridium botulinum Type A neurotoxin
 sequence"
 gene 9..1325
 /gene="botA"
 CDS 9..1325
 /gene="botA"
 /codon_start=1
 /product="botulinum neurotoxin serotype A Hc fragment"
 /protein_id="AA80610.1"
 /db_xref="GI:733428"
 /translation="MARLSTPEYIKNIINISLNRRESNHLIDLSRYAKINIGS
 KVFEDIENQIOLEFMESSKLEVLINAIIVNSMTEPSTSFWIRIPKYSISLNN
 EYIINMENNNSGMKVSILNYGELIWTLOETIKORVRYKVSQMINISDYINRMIVT
 ITNRNLSKIYINGRLIDOKPISNGINHASNNIKFKDGCRODFRIYIKYFNLFED
 KELNEKEIDVIDYNOSSNGILKDFMGDILOYKPYMLLYDPNRYVDVNVNGIGYM
 YLKGPRGSVHTTIYIYINSSLYRTRKRIYIKKYSAGKNDNTYRNDRYINWYKKNEXR
 LATNMAOGVEKTLISALEIPDGNLSQVYVMSKNDGIFNCKKMLQDNNNDIGFI
 GFHFNNAIKLVASNNYNNROIERSHFTLCSWEFLPVDGWERPL"
 BASE COUNT 401 a 342 c 249 g 346 t
 ORIGIN
 Query Match 98.6%; Score 1304.4; DB 12; Length 1338;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 12 gatgtctacaccttaacataacgaagaacatcatcatcactcactcgaacctg 71
 Db 20 GCTGTCTACCTTCACTCAATATCATCAAGAACATCATATCACTCCATCTGTGAACCTGGC 79
 QY 72 ctacgaatccaatcacctgtacgcgtgtctgcgtacacgttccaaatcaacatcgtgtc 131
 Db 80 CTACGAATCCAATCACCTGTATCGACCTGCTCGCTACGCTTCCAAATCAACATCGGTTG 139
 QY 132 taaagttaactcgtatccgacgcgaagaatcgaatcgaatcgaatcgaatcgaatcgtc 191
 Db 140 TAAAGTTAACTTCATCCGATCCGATCGACCAAGAAATCAATCAGCTGTTCATATCGAAATCTTC 199
 QY 192 caaatcgaagtatcctcgaagaatgtctatcgtatatacaactctatgtacgaagaactctc 251
 Db 200 CAAATATGGAATGTATCTGTGAAGATGCTATCGTATCAACTATATGACGAATAACTCTCTC 259
 QY 252 caactccttctgtatccgtatccgaaatcacttcaactcactcactcctcgtgaagaatga 311
 Db 260 CACCTCTTCTGTGATCCGATCCGGAATCTTCACTCATCTCTGTGAACATGATATA 319
 QY 312 caccatatacaactgcatacgaagaacaaatctgtgtggaagtatctctgaactcgaactg 371
 Db 320 CACCATATCACTGATGATGGAACAAATCTGTGTGGAAGTATCTGTGAACATACGCTGA 379
 QY 372 aatcatctgactctgcaggaacactcaggaatcaacaacagcgtgtgtatcacaatactc 431
 Db 380 AATCATCTGGACTGTGACGACACTCGAAGAAATCAACACGAGTGTATTTCAAAATATCTC 439
 QY 432 tcagaatgaacaactctcgtactacatcaatcgtcgtgatactcgtgtacacacaa 491
 Db 440 TCAATGATGCAACATCTGTGACTGATCAATCCGCTGATCTGTGTACATCAACAA 499
 QY 492 tgcgtctgaataactcgaataatcacaatcgaacgacctgtatctgacgaagaacagatctc 551
 Db 500 TCGTCTGAATATACCTCCAAATATCAATCAACAGCGCGCTGTATGACCAAGAACGATCTC 559
 QY 552 caatctggtgaacatcagcgtcttaataacatcatgttcaaaactggaagctgtgtgtga 611
 Db 560 CAATCTGGGTAACTCCACGCTTCTATTAATCATATGTTTAAACTGACCGCTTGTCTGTA 619
 QY 612 cactcaccgtctacatcgtgaatcaataactcgtgtcgtgaagaagaactggaaga 671
 Db 620 CACTCACCGCTACATCTGGAATCAATATCTTGTGACAAAGAACTGAAACGAA 679
 QY 672 agaaatcaagaacctgtacgaacaaccagtcacattctgtatccttgaaagactctctggg 731

Db 680 AGAATCAAGACCTGTACGACACACAGTCCAAATCTGTCTGTAAGACCTTGGGG 739
Oy 732 tgactaacctgagtagcaacaacccgtactacatctgaatctgtagatccgaacaata 791
Db 740 TGACTACCTGAGTAGCAACAACCGTACTACATCTGTAATCTGTACGATCCGAACAATA 799
Oy 792 cgttagcgtcaacaatgtagtaccgcggttacctgaatgtagtgaagtcgcggtgttc 851
Db 800 CGTTGACGTCAACATGTAGTATCCGCGTTACATGTACTGAAAGTCCGCGTGGTTC 859
Oy 852 tgttagtactcaacaatctactcctgaactctccctgtagcgtgtgtagtccaatcatcat 911
Db 860 TGTATACTACCAACATCTACTGAACTCTTCCCTGACCGTGTACCAAAATTCATCAT 919
Oy 912 caagaatacgcgtctgtgtaacaagaacaatactgttcgcgaacaatgtagtatacat 971
Db 920 CAAGAAATACCGCTGTGTGTAACAAGCAATATGCTTCGCAAAATGATCCTGTATCAT 979
Oy 972 caatgtgttagtlaagaacaagaataccgtctgtgtagtccaatgtcttcaggctgtgt 1031
Db 980 CAATGTTGTACTTAAGAACAAGATACCGTCTGGCTACCAATCTCTCAGCGTGGTGT 1039
Oy 1032 agaaaagatctgtctgtctgtgaaatcccgagcgtgtgtaactgtctcagtagtgt 1091
Db 1040 AGAAAGATCTTGTCTGTCTGTGAAATCCGAGCTTGTGTATCTGTCTGAGTAGTGT 1099
Oy 1092 aatgaatccaagaagaacccgggtatcaactaacaataatgcaaaatgtaactgtcagaaca 1151
Db 1100 AATGAATCCAAAGAACACACGAGGTATCACTAACAAATGCAATGTATCTGCGAGGACAA 1159
Oy 1152 caatgtagtagatacgtttcatcaggtttccaccagttcaacaatacgttaacaggt 1211
Db 1160 CAATGTGTACTTAAGAACAAGATACCGTCTGGCTACCAATCTCTCAGCGTGGTGT 1219
Oy 1212 tgcctcaactgtgtagaactcgtcagatcgaacagttcctcctcgcactcgtggtgtctgt 1271
Db 1220 TGTTCACAACTGTACATCATCTGATGCAATGCAATGCTTCTGCGACTGTGGGTCTCTGTG 1279
Oy 1272 ggaattcatcccggtgtgtagagcgtgtgtggtggaacgtccgtgtaa 1317
Db 1280 GGAATTCATCCCGGTGTGATGACGGTGTGGGTAACGTCCGCTGTAA 1325

RESULT 5
AR000030 1402 bp DNA linear PAT 04-DEC-1998
LOCUS AR000030
DEFINITION Sequence 25 from patent US 5736139.
ACCESSION AR000030
VERSION AR000030.1 GI:3962561
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 1402)
AUTHORS Kink,J.A., Thalley,B.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.
TITLE Treatment of Clostridium difficile induced disease
JOURNAL Patent: US 5736139-A 25 07-APR-1998;
FEATURES
Location/Qualifiers
source
BASE COUNT 420 a 360 c 260 g 362 t
ORIGIN

Query Match 98.6%; Score 1304.4; DB 6; Length 1402;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 12 gatctacactcactgaatacatcaagaacatataataaccatccatccgaaccgtcg 71
Db 84 GCTGTACTACTACTGATATCATCAAGAACATCATCAATATCTCCATCTGAACTGCGG 143

Oy 72 ctacgaatccaatcaactgtatcgaccgtgtctcgtctacagcttccaaatcaacatcggttc 131
Db 144 CTAGCAATCAATACCTGATTCGACCGTCTGGCTACGCTTCCAAATATCAACATCGGTTG 203
Oy 132 taaagttaactcgaatccgaatcgacaagaatcagatccagctgttcaatctggaatcttc 191
Db 204 TAAAGTTAACTTGATCCGATCGACAAAGATTCAGATTCACAGTCTTCATCTGAAATCTTC 263
Oy 192 caaatcgaagttaactcgtgaagaatgtagtcfatcaactcatctatgaagaactcttc 251
Db 264 CAAATCGAAGTTATCTGTGAAGATGCTATGATCAACTCTATGTAGCAAAATCTTCTC 323
Oy 252 caccctcctctgtatcccgtaatccgaataactccaactccatctcctgacaaatgaata 311
Db 324 CACCTCCTTGTGATCCGTTATCCCGAATATCTTCACTCCATCTCTCTGACAAATGAATA 383
Oy 312 caccatcaactcgtcgtgtaaaacaattctgtgtggaagtagtctctgtaactaggtga 371
Db 384 CACCATCATCACTGATGGAAGAAACAATCTGTTGGAAGTATCTGTCAATCACTACGTTGA 443
Oy 372 aatcatctgacctcgtgagaacactcaagaataccaacagcgtgtgtatcaaatctc 431
Db 444 AATCATCTGGACTCTTGACAGACACTCAGAAATCAACACCGTGTGTATTAATATCTC 503
Oy 432 tcagatgatacaactctcgtactacaatcaactcgtgatctcgttaccataccaaca 491
Db 504 TCAGATATTAACATCTCTGATCAATCAATCGCTGGATTTGTTACCACTCACACAA 563
Oy 492 tctgtgaatcaactccaanaactcaatcaacaaagcgtctgtatcgaccagaacagatctc 551
Db 564 TCGTGAATTAATCTCCAAATCTATCATCAACGCGCTGTGATGACAGAAACGATCTC 623
Oy 552 caatctggtgtaaatccacgccttctaataacatcagttcaactcgtgagcgtgtgtgta 611
Db 624 CAATCTGGGTAAATCCACGCTTCTTAATATCATCATGTTCAATCTGACGCTGTGTGTA 683
Oy 612 cactcaaccgtataactcgtgataactcaactcgttctgttgcgaagaagaactgaga 671
Db 684 CACTCACGCTTACTCATCTGATTCAAATCTTCAATCTGTTGACAAAGAACTGAAACGAAA 743
Oy 672 agaaatcaagaacctgttagaacaacagttccaattctgtgtagtctgaaagaactctggg 731
Db 744 AGAATCAAGACCTGTACGACAAACAGTCAATCTGTTGATCTGTAAGAACTTCTGGGG 803
Oy 732 tgactaacctgagtagcaacaacccgtactacatctgaatctgttagatccgaacaata 791
Db 804 TGACTACCTGACGATGACCAACCGTATCATGTGTAATCTGTAGATCCGAACAATA 863
Oy 792 cgttagcgtcaacaatgtagtaccgcggttacctgaatgtagtgaagtcgcggtgttc 851
Db 864 CGTTGACGTCAACATGTAGTATCCGCGTTACATGTACTGAAAGTCCGCGTGGTTC 923
Oy 852 tgttagtactcaacaatctactcctgaactctccctgtagcgtgtgtagtccaatcatcat 911
Db 924 TGTATACTACCAACATCTACTGAACTCTTCCCTGATCCGTTACCAAAATCTCATCAT 983
Oy 912 caagaatacgcgtctgtgtaacaagaacaatctgttcggaagaagaatgtagtatacat 971
Db 984 CAAGAAATACCGCTGTGTGTAACAAGCAATATGCTTCGCAAAATGATGTTATATCAT 1043
Oy 972 caatgtgttagtlaagaacaagaatacgcgtctgtgtagtccaatgtcttcaggctgtgt 1031
Db 1044 CAATGTTGTACTTAAGAACAAGATACCGTCTGGCTACCAATGCTTTCAGCGTGGTGT 1103
Oy 1032 agaaaagatctgtctgtctgtgaaatcccgagcgtgtgtagtctgtctcagtagtgt 1091
Db 1104 AGAAAGATCTTGTCTGTCTGTGAAATCCGCGAGTGTGAAATCTCTCAGGTAATCTGT 1163
Oy 1092 aatgaatccaagaagaacccgggtatcaactaacaataatgcaaaatgtaactgtcagaaca 1151
Db 1164 AATGAATCCAAAGAACACACGAGGTATCACTAACAAATGCAATGTATCTGAGGACAA 1223
Oy 1152 caatgtagtagatacgtttcatcaggtttccaccagttcaacaatacgttaactgt 1211

Db 1224 CAATGCTAACATATCGGTTTCATCGGTTTCACACAGTTCAACAATATCGCTAAACGTG 1283
 Oy 1212 tgcctccacacgtgacaaatcgtccagatcgaaacgttcctcctcgcactcgtggctcctg 1271
 Db 1284 TGCTTCCAACTGATGACATCTCGAGATCGAAGGTTCTCTGCACTCGGGGTTGCTCTTG 1343
 Oy 1272 ggaattcaccggcttgatgacggttggtggaacgtccgcctgtaa 1317
 Db 1344 GGAGTTCATCCCGGTTGATGACGGTTGGGGTGAACGTCCGCTGTA 1389

RESULT 6
 AR169141
 LOCUS AR169141 1402 bp DNA linear PAT 17-DEC-2001
 DEFINITION Sequence 25 from patent US 6290960.
 ACCESSION AR169141
 VERSION AR169141.1 GI:17906911
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1402)
 AUTHORS Kink,J.A., Thalley,B.S. and Stafford,D.C.
 TITLE Vaccine and antitoxin for the treatment of C. difficile disease
 JOURNAL Patent: US 6290960-A 25 18-SEP-2001;
 FEATURES location/Qualifiers
 source 1..1402
 BASE COUNT 420 a 360 c 260 g 362 t
 ORIGIN

Query Match 98.6%; Score 1304.4; DB 6; Length 1402;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 12 gatgtctaccccttcacatgatacatcaagaacatcatcaatacctccatccctcgtgaacgtg 71
 Db 84 GCTGTCTACCTTCACCTGAAATACATCAAGAACATCATCAATCATCTCCATCTCGAACCTCG 143
 Oy 72 ctagaatccaatcaacacgtcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 131
 Db 144 CTAGGAATCCAAATACCTGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 203
 Oy 132 taaagttaactcgtacgcgatcgcgaagaacatcagatccagctgttcaactcgtgaactcttc 191
 Db 204 TAAAGTTAACTTCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 263
 Oy 192 caaaatcgaaatctcctgaagaatgctatcgtatcaactctatgttgaagaactcttc 251
 Db 264 CAAATCGAAGTTATCCGAGAAATGCTATGATACAACTATGATGATGATGATGATGATGATGAT 323
 Oy 252 caccctcctcgtgacgtcgtatcccggaataactcaactccatcctcctcgtgaagaatgaata 311
 Db 324 CACCTCTCTTGAGATCCGTATCCGAAATCTTCAACTCCATCTCTCTGGAACATGATGA 383
 Oy 312 caccatcaaacgtacgagaaacaaatctcgtgtggaagaatctcctcgtgaactcgtgga 371
 Db 384 CACCATCATCAACTGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 443
 Oy 372 aatcatctgacgtcgtcgaagaactcaggaatacaaacacgctgttlatcaaatatc 431
 Db 444 AATCATCTGGACTCTGCGAGCACTCAAGGAATCAAAACGCGTGTGATTCATCAATACTC 503
 Oy 432 tcaatgatcaacaatctcgtatcatcaatcgtcgtggtggtcgttccatcacacaata 491
 Db 504 TCAGATGATCAACATCTCTGATCAATCAATCGCTGATCTCTGATCAATCAATCAATCAATCA 563
 Oy 492 tcgcttgataatacccaaaactcatcatcaagcgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 551
 Db 564 TCGTCTGATTAATCCAAATCTATCAATCAAGCGCGTGTGATTCGACCAAGAACCGATCTC 623

Oy 552 caatctggtgaacataccacgcttctaaatacatatgttcaaatcgtgacggttgcgtgta 611
 Db 624 CAATCTGGATTAATCCACCGCTTCTAATATCATATCTTCAAACTGACCGTGTGCTGA 683
 Oy 612 cactcaccgcttaactcctcgtgataataactcaatctcgttgcgaagaacttgaaacgaaa 671
 Db 684 CACTCACCGCTTACATCTGAGATCAATCTTCAATCTGTTGACCAAAAGACTGAAACGAAAA 743
 Oy 672 agaaatcaaaagacctgtacgacaaacagctcaatcttcgtatcctcgtgaagaactctgg 731
 Db 744 AGAATATCAAGACCTGTATGACACAAACGATCAATCTGATCTCTGAAACACTCTTGCGG 803
 Oy 732 tgactacctcgaatgaagaacacccgtactataatgtgaatctcgtgaatcgcgaacaata 791
 Db 804 TGACTTACTCGACATGACACAAACCGCTACTCATGCTGAATCTGTACGATCCGAACAATA 863
 Oy 792 cgttgaggttaaaatgtatggtatccggttgcgttgcgttgcgttgcgttgcgttgcgttgc 851
 Db 864 CGTTGACGCTCAACATATGATGATCCCGGTTACATGTACTGTAAGGTCGGCGTGGTTC 923
 Oy 852 tgtatgactacaacaactcaactcgtgaactcttcctcgttacggtgtgataccaatcatcat 911
 Db 924 TGTTATAGTACCAACATCTAATCTGAACTCTTCCCTGTACCGGTGATCAAAATTCATCAT 983
 Oy 912 caagaatacgcgtcgtgtaacaagaacatlatcgtcgaacaatgacgtgtatcatat 971
 Db 984 CAAGAATATACGCGCTGTTGAACAAGCAATATGTTGCGCAACATGATGTTGATCAT 1043
 Oy 972 caatggtgtgtaagaacaaagataccgttcgtgcacaaatgctcctaagctcgtggt 1031
 Db 1044 CAATGTTGATGATTAAGAACAAAGATACCGCTGTGGCTACCAATGCTTCTCAAGGCTGTGT 1103
 Oy 1032 agaaatgactctgtcgtcgtcgtgaaatcccggaactcgtgtgaatcgtcgtcgtcgtcgt 1091
 Db 1104 AGAAATGATCTTCTGCTGCTGGAATCCCGGAGCTGTGATATCTGCTCAGGATGTGT 1163
 Oy 1092 aatgaatccaagaacgacgaaggtatcactaacaatgcaaaatgaaatcgtcgaagcaaa 1151
 Db 1164 AATGAATCCAAAGAACACCAAGGATATCACTAACAAATGCAAAATGATTCGACGACAA 1223
 Oy 1152 caatggtgaagataggtttcatcgtttccaccagttcaacaatacgtcctaactggt 1211
 Db 1224 CAATGTTAGATATCGGTTTCATCGGTTTCCACCACTTAAACATATCCCTAACTGAT 1283
 Oy 1212 tgcctccaactggtgacaaatcgtcagatcgaagcttccctcgtcgcactcgtggttgcctct 1271
 Db 1284 TGCTTCAACGTGATCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1343
 Oy 1272 ggaattcaccggcttgatgacggttggtggaacgtccgcctgtaa 1317
 Db 1344 GGAGTTCATCCCGGTTGATGACGGTTGGGGTGAACGTCCGCTGTA 1389

RESULT 7
 AX036246
 LOCUS AX036246 1402 bp DNA linear PAT 16-NOV-2000
 DEFINITION Sequence 25 from Patent EP1041149.
 ACCESSION AX036246
 VERSION AX036246.1 GI:11225864
 KEYWORDS
 SOURCE synthetic construct.
 ORGANISM synthetic construct.
 REFERENCE 1 (bases 1 to 1402)
 AUTHORS Kink,J.A., Firca,J.R., Padhye,N.V., Thalley,B.S., Stafford,D.C. and Williams,J.A.
 TITLE Vaccine and antitoxin for treatment and prevention of C. Difficile disease
 JOURNAL Patent: EP 1041149-A 25 04-OCT-2000;
 FEATURES location/Qualifiers
 source 1..1402
 /organism="synthetic construct"

```

/db_xref="taxon:32630"
1..1389
/notes="Synthetic"
/codon_start=1
/translation="CAGC16479.1"
/db_xref="GI:11225865"
/translation="MGHHHHHHHHHHHSGHIEGRHMASMARLSTFEYIKIINTSI
LNRYESNHLIDLSRYASKINIGSKVNPIDKNOIOLFESSKIEVLILKNAIVNS
MYENFSTFWIRIPKYPFNSISLNNETIIMCMNNSGKSLWGEIYITLOPTEIK
ORVETYSQMINISDYINWIEVITNNRINNSKIYINDLIDOKPISLGNTHASN
IMFKDGCBDPRHYIMKYFNEPDKELNEKEIDLDNOSNGLDPMGDYLOYPK
YMLNLYDPKRYDVNVNNGIIRGYMLKPGSYMTNININSSLYRGTEIIRKVASG
NKDNIYRNNDRVINYVVKREYRLATNAGAVEILSLTEIPDGNLSQVYVMSK
NDGINTKCKMNLQDNNNDIGFIFGHOFNINIKLIVASNNYNOIERSSTLCCSNEF
IPVDDGKGEPRP"

BASE COUNT      420 a      360 c      260 g      362 t
ORIGIN
Query Match      98.6% Score 1304.4 DB 6 Length 1402:
Best Local Similarity 99.9% Pred. No. 0:
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0:

QY 12 gatgtcactctcaatgatacaagaacatcatcactcctcgaactcgcg 71
    |||||||
DB 84 GCTGTACTACTCTGATGATACATCAAGAACATCATCATCTCCATCCGAACTGCG 143

QY 72 ctacgaatccaatcactgatcgactgtctgctacgcttccaatcaacatcggttc 131
    |||||||
DB 144 CTACGAATCCAAATCAGCTGATGACCTGTCTGCTACGCTTCCAAATCAACATCGGTTTC 203

QY 132 taaagtaacttgatcgcgatcgacgaagaatcagatcagctgttcaatcgtgaactcttc 191
    |||||||
DB 204 TAAAGTTACTTCCATCCGATCGACACAGAAATCAGATCCAGCTGTTCATCTGGAAATCTTC 263

QY 192 caaaatcgaaatgtatccttgaagaatgtctatcgtatatacaactctatgtacgaagaactcttc 251
    |||||||
DB 264 CAAATGGAAGTGTATCCGAGAAATGCTATGATACACCTCTATGTAAGAAACTTCTC 323

QY 252 caactctcttgatcgtatcgtatcccgaaataacttcaactcactctcttgaaatgaata 311
    |||||||
DB 324 CACCTCTCTTGATTCGGATCCGTAATCTTCACTCCATCTCTCTGAAACAATGATA 383

QY 312 caccatcaatcaatcgatgagaagaacatctgtgttggaagaatctctggaactcgtgtga 371
    |||||||
DB 384 CACCATCATCACTGCATGCAATGAAAAATCTTGGTTGGAAATATCTCTGAACTACGGTGA 443

QY 372 aatcatctggaactctgacgaactcaggaatacaaacagcgtgtgttatcaatactc 431
    |||||||
DB 444 AATCATCTGACCTCTGCGAGACTCAGGAATCAACACAGCGTGTGTAATCAATATCTC 503

QY 432 tcaatgatcaaatctctctgactacataatcgtctgatatctgtttaccatccaataa 491
    |||||||
DB 504 TCAATGATGTAACATCTCTGACTACATCAACCGCTGATCTCTTACCATCACCAACAA 563

QY 492 tgcgtcgaataacccaataatctcatcaatcagcgcgtctgtatgaccagaagaacgatctc 551
    |||||||
DB 564 TCGCTGTAATACTCCAAAAATCTACATCAAGCGCGCTGTGATCAACCAAGAAACCGATCTC 623

QY 552 caatctgggttaacatcgaactcttaataacatcatgtttcaactggaagcgtgtgtcgtga 611
    |||||||
DB 624 CAATCTGGGTAAATCCACGCTTCTAATTAACATCATGTTCAACTGAGCGGTGTCGTGA 683

QY 612 cactcaccgctacatcttgatccaataacttcaatctgtttgacaagaagactggaacgaaa 671
    |||||||
DB 684 CACTCACCGCTACATCTGGATCAATCAATCTTCAATCTGTCCAAAGAAAGCAAGCAAAA 743

QY 672 agaatcaagaactgtatcgaacaccgctcactcttgatctgctgaaagaactcttgagg 731
    |||||||
DB 744 AGAATCAAAAGACTGTACGACAAACAGTCCATTTCTGTGTAAGAGACTTCTGGGG 803

QY 732 tgactacctgagtagcagaacacgctactacatgctgaatctgtacgatccgaaacata 791

```

```

DB 804 TGACTACTCTGACATGACACAAACCGTACTACATGCTGAATCTGTAGATCCGAACAAATA 863
    |||||||
QY 792 cgttgacgtcaacaatgtagtatccggtgttacaatgataaagtcgcgtgttc 851
    |||||||
DB 864 CGTTGAGGTCAACAAATGATAGTATCCCGGTTACATGATCTGAAAGGTCCGGTGGTTC 923

QY 852 tttatgactaccaacatctaccctgaactcttcctctgactcgtgtgataccaatctcat 911
    |||||||
DB 924 TGTATACTACCAACTTACTGACTGAACTTCCCTCTGTACCGTGTGACCAATTCATCAT 983

QY 912 caagaataacgctctgtgtaacaagaacaaatalcgtctgcgaacaatgataatcgtgtatcat 971
    |||||||
DB 984 CAAGAAATAGCGCTGCTGTAACAAGCAATATGCTTCCGAACAATGATGCTGTATACAT 1043

QY 972 caatgttgtagttaagaacaagaatacgtctgtgctacaaatgctcttcaggctgtgt 1031
    |||||||
DB 1044 CAATGTTGTAGTTAAGAACAAAGAAATACCGCTGGCTACCAATGCTTCCAGCTGGTGT 1103

QY 1032 agaaagatctgtctgtctcgtggaatcccggaactgtgtatctgtctcagtagtgt 1091
    |||||||
DB 1104 AGAAAGATCTGTCTCTCTGTGAAATCCCGAGCTTGTGTAATCTGTCTAGGTAGTTGT 1163

QY 1092 aatgaatccaagaacgacccaggtatcacttaacaaatgcaaaatgaaatctgcaggaacaa 1151
    |||||||
DB 1164 AATGAATCCAAAGAACAGACAGGATATCACTAACAAATGCAAAATGAAATCTGCAGAGCAA 1223

QY 1152 caatgtgaacgatatcgtgttcatcgtgttccaccagttcaacaatatcgtctaactgtc 1211
    |||||||
DB 1224 CAATGTTAAGAAATATCGGTTTCACTCGGTTTCCACCACTTCAACAAATATCCCTAACTGGT 1283

QY 1212 tgcctcacaactgtgtacaaatcgtacagatcgaacttcctcgcgaactcgtgtgtgtctgt 1271
    |||||||
DB 1284 TGCTTCAACTGCTGACAAATGTCAGATCGAACGTTCTCTCGCACTCTGGGTTGCTTGTG 1343

QY 1272 ggaatcatcccggtgtgatgacggtgttgagggtgaacgctgcgtgttaa 1317
    |||||||
DB 1344 GGAATTCATCCCGGTTGATGACGGTGGGTGAACGTCGCTGTAA 1389

RESULT 8
CBOTAG
LOCUS      CBOTAG      4292 bp      DNA      linear      BCT 12-SEP-1993
DEFINITION      Clostridium botulinum botA gene for type A neurotoxin.
ACCESSION      X52066.1
VERSION      X52066.1 GI:40381
KEYWORDS      botA gene; neurotoxin; secreted protein.
SOURCE      Clostridium botulinum.
ORGANISM      Clostridium botulinum.
REFERENCE      1 (bases 1 to 4292)
AUTHORS      Minton,N.P.
TITLES      Direct Submission
JOURNAL      Submitted (08-JAN-1990) Minton N.P., PHLS Centre for Applied
REFERENCE      Microbiology & Research, Molecular Genetics Group, Division of
                Biotechnology, Porton Down, Salisbury SP4 0UG Wiltshire, U K
                2 (bases 1 to 4292)
AUTHORS      Thompson,D.E., Brehm,J.K., Oultram,J.D., Swinfield,T.J.,
                Shone,C.C., Atkinson,T., Melling,J. and Minton,N.P.
TITLES      The complete amino acid sequence of the Clostridium botulinum type
                A neurotoxin, deduced by nucleotide sequence analysis of the
                encoding gene
JOURNAL      Eur. J. Biochem. 189 (1), 73-81 (1990)
FEATURES
MEDLINE      90235864
SOURCE      Location/Qualifiers
                1..4292
                /organism="Clostridium botulinum"
                /strain="Sub sp. type A, NCTC2916"
                /db_xref="taxon:1491"
                16..64
                misc_feature
                promoter
                19..24
                /note="dyad symmetry"

```

```

promoter      /note="-35 region"
              41..46
              /note="-10 region"
RBS           63..68
              /note="ribosome binding site"
              77..3967
              /note="type A neurotoxin (AA 1-1296)"
              /codon_start=1
              /transl_table=11
              /protein_id="CAA36289.1"
              /db_xref="GI:40382"
              /db_xref="SWISS-PROT:P10845"
              /translation="MDFVNRKQFNTKDVNCGVDIAYIKIPNVGQMPYKAKIKINKTW
              IPERDFTNEEDDLNPPPEAKQVPVSYDSTYLSIDNEDNKLKGTLEKLEISD
              LGRMLLTSIVRGIPFMGSGTIDLEKVIDNCINVLOPDGYSSEELNLIIGPSADI
              IOECKSGFGEVNLNPNNGYSTGYIREFSDPFEESELEVTNPLGLGKATPDA
              VTLAHELIGHGHLGYGAIINPNVFKYNTAAAYEMSGLEYSPELIRPGGDAKFTDS
              LOENPERLYYNNFKDIASTLNKAKSIVGTASLOTKRNFKKKYLISEPTSKFSD
              KLNFDKLYKMLTEIYEDNFVKKFVLRNTYINFDKAVKINIVPVNTIYDGEVL
              RNTNLANFGQMTLEINNNMFTKLNFTGLPEFKLCVAGIITSKSLDKGYKNAL
              NDLCIKVNNMDFEFSPEDNFTNDLNGEITSDTNEAEENISDLIOQYLYTFE
              DNENPENSISLSDIIGOLEMPTERFNGKRYELDKTMTFHYLRAROEFGKSTI
              ALTNSVEALNPNRVTFFPSDYPVKVKNATPAAATGVEQLOLYDPTDPESEVST
              DKADITIIIPYIGPALNIGNMLTKDFPGALIFSGAVILLEPTELAIPIVLTFAIV
              SYIANRVLIVQITIDNALSKRNEKWEYKIVINWLAKVMTQIDLIRKKKEALEMDA
              EATKALINYOQYTEEEKNNINFINIDLSKKNESINKAMININFLNCGSYLKN
              SMIPYGVKRLDEPDASLDALRYIDNRTGLIGOVRLDKVNTLSTDIIPOLSKY
              VDNORLSTFEYIKNIINTSIILNRESNHLIDLSYASKINSNVEYTIINCMENS
              OLFNLSSEKTEVLKNAIYVNSWENSTSPFMRIPKPYEISINNEYTIIINCMENS
              GKRYSLATGEIITITLDDTOEIKORVYFKYSOMINISDIIRKAFVITTNRLNSKIY
              INGRIDOKPISLNGNTHASNNIMFKIDGCRDHRHYIWKIFNLPKLENEKIKIDY
              DNOSNGSLDKDFMGDYLYDKPYKMLLYDPNRYVDNNGVINGMYLKPQSGSVMT
              NITLNSLSLYRGTEFIKKYASGKNKDINVRNDDIYIVVYKNEYRLATNAAGSEK
              ILALEIPDGNLSOVVYVMSKNDQITNCKNMLDQNNNDIGIFGHQFNMIKLYV
              ASMWNRQIERSSRTLGCSNFIIPVDGMEERL"
misc_feature  4042..4087
              /note="dyad symmetry"
              4104..4136
              /note="dyad symmetry"
misc_feature  4158..4188
              /note="dyad symmetry"
BASE COUNT   1738 a      423 c      668 g      1463 t
ORIGIN
Query Match      55.0%; Score 727.6; DB 1; Length 4292;
Best Local Similarity 72.2%; Pred. No. 7.9e-177;
Matches 946; Conservative 0; Mismatches 364; Indels 0; Gaps 0;

```

```

QY 374 tcatctggactctgcaggacactcgaagaatcaaacgctgtgtatctaaactctc 433
DB 3024 TAATCTGAGCTTTTACAGGATATCTCAGGAATAAACAAGAGTAGTTTTAAATACAGTC 3083
QY 434 agatgatcaacatctctgactataactcgcgtggactctcgttaccatccaacaatc 493
DB 3084 AAATGATTAATATATATGAGATTATTAACAGATGAGATTGTTGTAACATCATCTAATAATA 3143
QY 494 gtcgtgaataactccaataactacatacaacgcgcgtctgataccgaacgcagatccca 553
DB 3144 GATTAAATTAATCTCAAAATTTATATATAATGAGATTAAATAGATCAAAACCAATTTCAA 3203
QY 554 atctgggtaacatccacgcgtcttaataacatcaltgatacaactcggacgtgtcgtgaca 613
DB 3204 ATTATAGCTAATATTCACAGCTAGTAATATATATATATATATATATAGATGCTGAGAGATA 3263
QY 614 ctacccgctacatctgatacaaatcactcaactcgtctgcgaagaagacgtgaagaagaag 673
DB 3264 CACATGATATATTTGGATTAATATTTTATCTTTTGTATAGGAATTAATGAAAGAAAG 3323
QY 674 aaatcaaaagacctgacgacaaccagctccaatctgtatccctgaaagactctggggtg 733
DB 3324 AAATCAAAAGATTTTATGATATATCAATCAATTCAGGTATTTTAAAGACTTTTGGGCTG 3383
QY 734 actaccctgcagtaacgacaacgcgtactacatgctgaaatcgtacgatacgcgaacaatacg 793
DB 3384 ATTATTTACAAATATGATTAACCACTACTATATGTTAAATTTATATGATCCAAATAAATATG 3443
QY 794 ttgacgcgaacaatgtagtataccggtgtacatgtaaccctgaaaggtccgcgtgtcttg 853
DB 3444 TCGATGTAATTAATAGTAGTATTAAGAGTTATATGATCTTAAAGGGCTGAGAGTAGCG 3503
QY 854 tttagactacaacactactacactcgtgaacctccctccgtcgtgtgataccaaatcacaatca 913
DB 3504 TAATGACTACAAACATTTATTTATTAATTCAGAGTTTGTATAGGGGACAAATTTATATATA 3563
QY 914 agaaatacgcgtctgtaacaagaacaaatcgttcgcaacaatgatacgtgtatataca 973
DB 3564 AAAATATATGCTTCTCGAAATAAAGATATATTTGTGAATAATATGATCCGTATATATTA 3623
QY 974 atgttgtagttaaagaacaagaataaccgtctgcttaccaaatgcttcctcaagcgtgtgtag 1033
DB 3624 ATGTAGTAGTAAATAAATAAAGATATATAGTTAGCTACTAATGCAATCAGAGGAGGCGTAG 3683
QY 1034 aaaagatctgctcgtcctcgtgaatcccggaacgtctgtgtaactgctcgaagtggtgtaa 1093
DB 3684 AAAAATATAGTAAGTGCATTTGAATAATACCTGATGAGCAAAATCAAGTAGTAGTAA 3743
QY 1094 tgaataccaaagaacgcagcaggtatcaactacaataatgcaaaatgaaatcgtcaggaacaca 1153
DB 3744 TGAAGTCAAAATAATGATCAAGGAATTAACAATAATATGCAAAATGCAAAATTTTCAAGATATA 3803
QY 1154 atgttgacgataatcgtgttcatcgtgttcacacgaagttcaacaatatcgtccaactcgtgtg 1213
DB 3804 ATGGGAATGATATATAGCTTTATAGATTTTCATGAGTTTAAATATATAGTAAACTAGTAG 3863
QY 1214 ctctcaactgttgtaacaatcgtcgaatcgaacgtcttcctcgtcaactcgtgtgtcgtctg 1273
DB 3864 CAAATATTTGAGTAAATATAGCAAAATAGAAAGATCTAGTAGAGCTTTGGTGCATAGGG 3923
QY 1274 agttacatccggtgtgatacgcgtgtgggtgtaaacgctcgtcgtgaagaatc 1323
DB 3924 AATTTATTTCTGTGATGATGATGAGGAGAAAGGCCACTGTATTAATATC 3973

```

RESULT 9
 CLONEDUR 4835 bp DNA linear BCT 26-APR-1993
 LOCUS C.bolulinum neurotoxin gene, complete cds.
 DEFINITION M30196
 ACCESSION M30196
 VERSION M30196.1 GI:144864
 KEYWORDS neurotoxin.
 SOURCE C.bolulinum (strain 62A, subtype A) DNA.

D	2648	ATGAAATGATTCATTAAATGACTTATCTAGTATGCAATCAAAAATAAATATGTGTGTA	2707
O	134	aagtttaacttcgataccgatacgcagaataatcagataccgctgttcaacttgyaatcttcca	193
D	2708	AAGTAAATTTTGTGATCCAAATAGATATAAAAACAAATTCATTTATTTATTTAAAGTGTGTA	2767
O	194	aaatcgaagttatccggaagaatgcatctgtatatacaactctatctgacgaaactttcca	253
D	2768	AAATGAGGTAATTTTAAAAAATGCAATGTCTATATATATATATATATGATATGATATTTTACTA	2827
O	254	ccctcttcgataccgatalcccgaaatacttcaacttcaactctctctgtacaaatgaataca	313
D	2828	CTAGCTTTTGGATTAAGAAATTTCCATAGTATTTTAAACAGTATAGTCTAAATATGAAATTA	2887
O	314	ccatctatcaactcattggaagaacaatttcgtttggaagaatctctctgtacaggttgaa	373
D	2888	CATTAATTAATTTTATGAGAAAATAATTCAGAGATGGAAATATCACTTATATATGATGTA	2947
O	374	tcaatctgacatcgcaggaacactcagaataatcaaacagctgttattcaataatctc	433
D	2948	TAACTCTGACTTTACAGGATACACAGAAATTAACAAAGATATTTTAAATATCAAGTC	3007
O	434	agatgatcaacatctctgtactacataatcgcgtgtatctgttaccataccacaatc	493
D	3008	AAATGATTAATATATAGATTATATTAACAGATGGATTTTGTCTATCACTATATATA	3067
O	494	gtctgaataactcgaataatctatatacagcgccgtctgtacgcaggaacagatctcca	553
D	3068	GATTAAATACCTTAAATTTATATTAATGGAAGATTAATAGATCAAAAACCAATTTCA	3127
O	554	atctggttaacatccacgccttctaataacatgltcaaacatgacggttgcgtgaca	613
D	3128	ATTATAGTATATTCATGCTAGTAATTAATATATATGTTTAAATATGATGGTGTGACAGTA	3187
O	614	ctacccgctacatctgtatcaaatcttcaactctgttcgacaagaactgacgaanaag	673
D	3188	CACATATATATTTTGGATTAATAATATTTTAATCTTTTGTATAGAGAAATTAATGAAANAAG	3247
O	674	aaatcaagaacccgttgcagacaacacgatacttccgttgcataccgaaagactcttgggtg	733
D	3248	AAATCAAGAATTTATATATGATATATCAATCAAAATTCAGGATTTTAAAGACTTTGGGGTG	3307
O	734	actaactgcagtcgcagacaacgcgtactacatgctgtaactctgtagatccgacaacaatcg	793
D	3308	ATTATTTTACATATGATTAACCATACTATATATGTTAAATTTATATATCAATAAATATG	3367
O	794	tttagcgtcaacaatgttagatatacccgcggttataatctatccatgaagaagtcgcgltgtc	853
D	3368	TCCATGTAATTAATGTATGATGATTAGAGGTTATATGTATCTTAAAGGCCCTTAGAGGTACCG	3427
O	854	ttatgataccaacatctactctgtaactcttccctcttaccgttgcataatctatcatca	913
D	3428	TAAATGACTTAACAAACATTTATTTAAATTCAAAGTTTGATATGGGGGCAAAAATTTATTTAA	3487
O	914	agaataatcggtctgttgaacaaggaacaatatogtttcgcacaacatgatatgttatacatca	973
D	3488	AAAAATATGCTCTCGAAATTAAGATTAATTTGTTATGAAATATATATATGCTGTATATATTA	3547
O	974	atgtgttagttlaagacaagaatatacgcgtctggtcccaatctcttccagctgtgttag	1033
D	3548	ATGTAGTAGTTTAAATAAATAAGATATAGGTTAGGCTATCTATGCTATCACAGGAGGCGTAG	3607
O	1034	aaaaagatctgtctgtcttgcgttgaataatccgcgagctgtgttataatctgtctcaagttagttga	1093
D	3608	AAAAAATATCTATGACTTATGAAATATCTGATGATAGGAATTTAAGTCAAGATATATAGTAA	3667
O	1094	tgaatccaagaacgcagcaggtgatactactcaacaatgycacaatgtatctgcagggaca	1153
D	3668	TGAAGTCAAAAATATGATCAAGGATATTAACAATTAATAGCAAAATGCAATTTACAAAGTATATA	3727
O	1154	atggtaacgataatcggtttcaatcggtttccacccagttcaacaatatcgctaaactgtgtg	1213

Accession	Sequence	Position
Db 3728	ATGGAGTGAATTAAGCGTTTATAGCATTCATCAGTTTAATTAATAGCTAAACATAGAG	3788
Oy 1214	cttccaaactgtgacaaatcgtccagatgaacagttccctccagactcgggtgttccttggg	12737
Db 3788	CAAGTAAATGGTAAATAGACAAATAGAAATGAAAGATCTAGTAGAGCTTTGGGTTCTCATGG	3844
Oy 1274	agttcatcccggtgtgatgacagcgttgggtgtgacgctccgctgttaa	1317
Db 3848	AATTATATCCGTAGATGATGATGGGGAAGAAAGGCGACCTGTAA	3891

RESULT	12			
LOCUS	AX036248			
DEFINITION	AX036248	3891 bp	DNA	linear
ACCESSION	Sequence 27 from Patent EP1041149.			PAT 16-NOV-2000
VERSION	AX036248			
KEYWORDS	AX036248.1	GI:11225866		
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS	1 (bases 1 to 3891)			
	Kink,J.A., Flicra,J.R., Padhye,N.V., Thallej,B.S., Stafford,D.C. and Williams,J.A.			
TITLE	Vaccine and antitoxin for treatment and prevention of c. Difficile disease			
JOURNAL	Patent: EP 1041149-A	27	04-OCT-2000;	

```

/organism="synthetic construct"
/db_xref="taxon:32630"
1. .3891
CDS

```

BASE COUNT 1580 a 392 c 629 g 1230 t
ORIGIN

Query Match	54.9%	Score 726.4	DB: 6	Length 3891
Best Local Similarity	72.3%	Pred. 1.6e-116		
Matches 943	Conservative 0	Mismatches 361	Indels 0	Gaps 0
Oy	14	tgctactccttcacgaaatcacatcaagaacatcatcaactctcatcctgaactgcgc	73	
Db	2588	tatctacatttactgtaaatatatttaagaattattttttaaacctctcttatggaatttaaacat	2647	

LOCUS	CBNT0XA	4067 bp	DNA	linear	BCT 24-JUN-1994
DEFINITION	C. botulinum gene for infant neurotoxin type A.				
ACCESSION	X73423				
VERSION	X73423.1 GI:507070				
KEYWORDS	botulinum neurotoxin; botulinum neurotoxin type A.				
SOURCE	Clostridium botulinum.				
ORGANISM	Clostridium botulinum Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.				
REFERENCE	1 (bases 1 to 4067) Willems A., East A.K., Lawson P.A. and Collins M.D. Sequence of the gene coding for the neurotoxin of Clostridium botulinum type A associated with infant botulism: comparison with other clostridial neurotoxins Res. Microbiol. 144 (7), 547-556 (1993)				
JOURNAL	2 (bases 1 to 4067) Willems MCG, A. Direct Submission Submitted (17-JUN-1993) A. Willems MCG, AFRC Institute of Food Research, Dept of Microbiology, Reading Laboratory, Earley Gate, Whiteknights Road, Reading, RG6 2EF, UK				
TITLE	Location/Qualifiers				
FEATURES	1..4067				
SOURCE	/organism="Clostridium botulinum"				
	/strain="Kyoto-F"				
	/db_xref="taxon:1491"				
RBS	83..88				
CDS	97..3987 /codon_start=1 /transl_table=1 /product="botulinum neurotoxin type A" /protein_id="CA51824.1" /db_xref="GI:507071" /db_xref="SPTRMBL:045894" /translation="MPFVNKQKDPNGVDIAVYIKIPNAGMOPVAKFIHNKIMW IPEDPTFNPREGDINPPEAKQVPVSYDSTYISTDEKONYLKVTKLEPRISTST LGRMLLSTIVAGIIPFWGSGTIDTELKVIIDTCINIDPDSRSELELVIIIGPSAD IQFCEKSGHDVNLTRNGYSTQYIRSPDPTGFEELEVDNIPPLGAKFADFM VTLAEHLHAHRLYGLAINDNRFKVTNKAIVYEMSGLEVSFEELRTSGHDAKID LOENEFRLYYNKRFDVASTLNKAKSIIGTASLOYMKNVREKILSEDSGRSVY KLRDKLYKMLTELYTDENFVNFPRKVIYNRKVIYNKDPKVRINIYDENVYIKQGS KCALSTVNMDELFPSPSEDNFTNDLDRKEETTAOTNIEAAEENISLDLIQYYLTFD NDPENISIEENISDIIIGOLEPMPNIEEPFGKXYELDKYMFHLRAQEFHDSR ILNESAEBALLKPNVAATFESSKYVKINKVAEAFPMNADELYDPTDENVTYTN DKINDITIVPYIGPALNIGMLSKGEVEVLIITGVAMLEPIPEVALPYEGPAIIV SYIANKVLVOTINNALSKRNEKDEYKTYITVMMLKAVNTQIDLRKKKALENQ EATNAIINYOYNQYTEEKNININIDLSKRLENSISAMININIKFLDQCSVLENQ SMIPYAVKRLKDPASVADVLLKTYINRGTVLQVLDLKEVNNTLSADIPFOLSK VDNKLSTFTEPEYIKNIYNTSILSIYKKDLIDLSRGAKINIGDRVYISIDNQ KLINLESSTIEVILKNAIIVNSVMENSTPWINIPYEFKINININEVTIINCLENN GKMSLVNTEGELIYTLQDNKONTQVRYKRSQVMNIDYINRWIYVTTNRNLRKSKYI INGULIDQKPLISLGNIHASKIMFKLDGCDPRRYIMKIFNLFDELNKEIKDL DSQNSLKDMPGNYLQYDKPYALMLFDNKKVVDVNNIGIRGMYLKGPSGVVYTI NIYNSNLTGEGTKFIKKRYASGNDNIRANDRYIINVVANKRYELATNKGSGVEE ILSLNLETIDYGNISQVYVMSKDDOGIRNCKMMLQDNNGNDIGFIGPHLDNIKLV ASNNYNNQYKASASTFGCGSMEFTIYVDQGWESSL"				
misc_feature	1384..1386				

OY 378 ctggaactctgcaggaactcaggaatcaaacagcgtgtgtatccaactccaat 437
|||||
DB 404 CTGGACTCTGGCAGGACTGCTGTGTAACACACGAAACTGGTTTCAACTACACTCAGAT 463
|||||
OY 438 gatacaatctctgaatacaatcgcgtggaatctcgttaccatcaaccaatcgtct 497
|||||
DB 464 GATCTCTATCTCTGACTACATTAATAATGATCTTCGTTACTATCACTAACACCGTCT 523
|||||
OY 498 gaatactccaatactcaatcaacgagcgtgtgtatcgacagaaacgcatctccaatct 557
|||||
DB 524 GGGTACTCTCGTATCTATCAATCAACGGTAACCTGATCGATGAATAATCTATCTTAACT 583
|||||
OY 558 ggttaacatccacgcttccataatcaatcatgttccaacgaggtgtgtgcgtgacatca 617
|||||
DB 584 GGGTACATCCACGTTTCTGACAAACCTCTTCAAAATCGTTGGTGGACAGACAC--- 640
|||||
OY 618 ccggtacatctggaataaatactcaatctgttcgacaaagaactgaacgaaanaaat 677
|||||
DB 641 GCGTTACGTTGGTATCCGTTACTCAAAAGTTTGCACACTGAACCTGGGTAAACTGAAT 700
|||||
OY 678 caaagactgtacgacaaacgcatcaatctgtgtatccgaaagacttctgggtgacta 737
|||||
DB 701 GGAACCTCTGTACTCTGACGAACCGACCCGCTATCTGAAAGACTTCTGGGGTAACTA 760
|||||
OY 738 cctgcaatgcagaaacacgcatcaatctgtgtatcgatccgaacaataatcgtga 797
|||||
DB 761 CCTGCTGTACACAAACGTTACTACTGCTGACACTGCTCCGACTGACAAATCTATCAC 820
|||||
OY 798 cgtcaacaatgtagatccgaggttaacatgtacatgaagaagtcgaggtgtgtat 857
|||||
DB 821 TCAGAACTCT-----ACCTCCTGAACATCAACACGACGCGGTGTATCA 868
|||||
OY 858 gactaccaacatctaccgaactctccctgtacccgtgtgtaaccaatctatcaaga 917
|||||
DB 869 GAAACCTATATCTCTCTATACACTGCTGTGTAACACTGCTGTGAAGTATCATCCGTAA 928
|||||
OY 918 atacggtct-----gtaacaagaacaatactcgttcgaacaatgatcgttata 968
|||||
DB 929 AAACGGTTTACTGACATCTCTAACACTGACACTGACACTGCTGTAACGTAACGCTTA 988
|||||
OY 969 caataatgtgtatgaataaagaataacggtctgtgctaccatgtctcgaagctg 1028
|||||
DB 989 CATCAACGTTGTTGACCGGTGGAATACCGTCTGTACGCTGACATCTATCCGTAA 1048
|||||
OY 1029 tctagaanaagatctgtctgcctctgaaatcccgagctgtgtaa---ctgtctcag 1085
|||||
DB 1049 ACCGGAATAATCATCAAACTGATCCGTACTTCTTAACACTCTTCTGGGTAGAT 1108
|||||
OY 1086 agtltgaataatcaagaacgacagggatcaactaacaatgcaaatgaatctga 1145
|||||
DB 1109 CATCGTTATGACTC-----GATCGGTAACTGCACTATGAACTTCCA 1153
|||||
OY 1146 ggaacaacaatggttaacagatatcgtgttcatcgtgttccacacggttcaacaatacgctaa 1205
|||||
DB 1154 GAACAACAACGGGTGATACGCTGCTGCTGGGTTTCCACTTAACACAC----- 1202
|||||
OY 1206 actgtgtgttccaaactgtgttaacatcgttcgaatcgaagcttccctcgcactcgtgtg 1265
|||||
DB 1203 -CTGGTGGTCTTCTTCTGTGTAACACAAACATCCGTAATAAACACTTCTTCTAACGGTGG 1261
|||||
OY 1266 ctcttggaggttcatcccggttgaatgaacggttgggtgaa 1305
|||||
DB 1262 CTTCTGCTCTTCACTCTAAGAACACGCGTGGCAGGAA 1301
|||||

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 21:42:55 : Search time 631.73 Seconds
(without alignments)
3595.647 Million cell updates/sec

Title: US-09-611-419A-3
Perfect score: 1323
Sequence: 1 gaattcgaaacgagtgtctac.....aacgtcgcgtctgaagaattc 1323

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq_032802:*

1:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1323	100.0	1323	22	AA54483	Botulism toxin hea
2	1310.4	99.0	1332	22	AA54482	Botulism toxin hea
3	1310	99.0	1326	22	AA54484	Botulism toxin hea
4	1304.4	98.6	1330	17	AAT29245	Type A neurotoxin
5	1304.4	98.6	1330	19	AAV30571	Clostridium botuli
6	1304.4	98.6	1338	21	AA287212	DNA encoding synth
7	1304.4	98.6	1351	19	AAV30576	Clostridium botuli
8	1304.4	98.6	1402	17	AAT29246	Type A neurotoxin
9	1304.4	98.6	1402	19	AAV30572	Clostridium botuli

10	1250	94.5	1332	22	AA54588	Sequence encoding
11	727.6	55.0	1546	19	AAV30575	Clostridium botuli
12	727.6	55.0	4835	21	AA64582	BoNT/A neurotoxin
13	726.4	54.9	1317	21	AA287220	DNA encoding nativ
14	726.4	54.9	2532	21	AA287218	DNA encoding nativ
15	726.4	54.9	3891	17	AAT29244	C. botulinum type
16	687.4	52.0	702	21	AA287221	DNA encoding BoNTA
17	619	46.8	621	21	AA287222	DNA encoding BoNTA
18	421.6	31.9	1313	18	AAT48101	Immunogenic type F
19	421.6	31.9	1314	22	AA54499	Botulism toxin hea
20	401.2	30.3	1317	21	AA287216	DNA encoding synth
21	399	30.2	1317	22	AA54490	Botulism toxin hea
22	371.8	28.1	1917	21	AA287219	DNA encoding nativ
23	289.4	21.9	1347	21	AA287217	DNA encoding synth
24	289.4	21.9	1368	22	AA54491	Botulism toxin hea
25	264.8	20.0	1347	22	AA54589	Sequence encoding
26	263.2	19.9	1278	22	AA54489	Botulism toxin hea
27	263.2	19.9	1400	21	AA287215	DNA encoding synth
28	263.2	19.9	1400	22	AA54488	Botulism toxin hea
29	263	19.9	1341	21	AA287213	DNA encoding synth
30	263	19.9	1341	22	AA54485	Botulism toxin hea
31	227	17.2	1293	18	AAT48100	Immunogenic type F
32	203.4	15.4	1374	22	AA54487	Botulism toxin hea
33	194.2	14.7	1472	19	AAV30585	Clostridium botuli
34	194.2	14.7	4017	22	AA58862	C botulinum BoNT/E
35	191	14.4	1463	19	AAV30584	Clostridium botuli
36	180.8	13.7	1460	19	AAV30593	Clostridium botuli
37	169.4	12.8	1371	21	AA287214	DNA encoding synth
38	169.4	12.8	1371	22	AA54486	Botulism toxin hea
39	166.4	12.6	1359	12	AA012121	Synthetic tetanus
40	166.4	12.6	3712	16	AA097490	Intermediate plasm
41	166.4	12.6	3754	15	AA057880	Intermediate plasm
42	166.4	12.6	3754	16	AA085420	Plasmid pREC1. S
43	166.4	12.6	3769	15	AA057881	Intermediate plasm
44	166.4	12.6	3769	16	AA097492	Plasmid pREC2. N
45	166.4	12.6	4366	16	AA085424	Plasmid pREC3-P28

ALIGNMENTS

RESULT 1	
ID	AA54483
AA54483	standard; DNA; 1323 BP.
AC	AA54483:
XX	
DT	11-APR-2001 (first entry)
XX	
DE	Botulism toxin heavy chain C-terminal coding sequence (serotype A).
XX	
KW	Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW	recombinant vector; antigen; Immune response; vaccine; bacterium;
KW	Infection; ds.
XX	
OS	Synthetic.
OS	Clostridium botulinum.
XX	
FH	Key
FH	Location/Qualifiers
FT	13.1317
FT	*tag= a
FT	/product= H_C peptide fragment
XX	
PN	MO200067700-A2.
XX	
PD	16-NOV-2000.
XX	
PF	12-MAY-2000; 2000MO-US12890.
XX	
PR	12-MAY-1999; 99US-0133865.
PR	12-MAY-1999; 99US-0133866.
PR	12-MAY-1999; 99US-0133867.
PR	12-MAY-1999; 99US-0133868.

Db 1281 ggaatccatcccgctgacgacgctgagtggaacgctccgctgaagaattc 1332

RESULT 3

AAA54484
ID AAA54484 standard; DNA; 1326 BP.

XX AAA54484;

XX 11-APR-2001 (first entry)

DE Botulinum toxin heavy chain C-terminal coding sequence (serotype A).

XX Botulinum toxin; neurotoxin; heavy chain; recombinant expression;

KM recombinant vector; antigen; immune response; vaccine; bacterium;

XX infection; ds.

OS Synthetic.

XX Clostridium botulinum.

FT Key Location/Qualifiers

FT CDS 13..1320

XX /tag= a

XX /product= H_C peptide fragment

XX WO200067700-A2.

XX 16-NOV-2000.

XX 12-MAY-2000; 2000MO-US12890.

XX 12-MAY-1999; 99US-0133865.

XX 12-MAY-1999; 99US-0133866.

XX 12-MAY-1999; 99US-0133867.

XX 12-MAY-1999; 99US-0133868.

XX 12-MAY-1999; 99US-0133869.

XX 12-MAY-1999; 99US-0133873.

XX 29-JUL-1999; 99US-0146192.

XX (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.

XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;

XX WPI; 2001-016048/02.

XX P-PSDB; AAB04050.

XX New nucleic acids encoding the carboxy- or amino-terminal portions of

XX the heavy chain of botulinum neurotoxin of serotype A-G, useful as

XX vaccine against botulinism

XX Disclosure; Fig 3a; 73pp; English.

XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide

XX chain and then posttranslationally nicked, forming a dichain which

XX consisting of a 100 kDa heavy chain and a 50 kDa light chain which

XX remain linked by a disulfide bond. Nucleic acids encoding the

XX carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy

XX chain of botulinum neurotoxin (BoNT) can be used in recombinant

XX expression vectors and expressed in transformed cells to produce

XX peptide antigens useful for eliciting an immune response to give

XX protective immunity against botulinum neurotoxin, which causes

XX botulism. The nucleic acids are expressible in a recombinant

XX organisms such as Escherichia coli or Pichia pastoris. The use

XX of recombinant nucleic acids are advantageous since it eliminates

XX the need to culture large quantities of hazardous toxin-producing

XX bacterium. Production yield from the genetically engineered product

XX is also high and cost of production is lower. The nucleic acids can

XX be derived from Clostridium botulinum serotypes A-G.

XX Sequence 1326 BP; 404 A; 336 C; 243 G; 343 T; 0 other;

Query Match 99.0%; Score 1310; DB 22; Length 1326;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1323; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 gaattcgaaacgagt---ttacccttcacgtgaataatcatcaagaacatcatcattcc 57
Db 1 gaattcgaaacgagtgccttcacgtgaataatcatcaagaacatcatcattcc 60
QY 58 atcttgaaacctgcgtacgacatcccaatccatccatccatccatccatccatccatcc 117
Db 61 atcttgaaacctgcgtacgacatcccaatccatccatccatccatccatccatccatcc 120
QY 118 atcaacatcggtcttaagttaacttcgatccgatccgacagaatacagatccagctgttc 177
Db 121 atcaacatcggtcttaagttaacttcgatccgatccgacagaatacagatccagctgttc 180
QY 178 aatctggaatctcccaaaatcgaaagttaatccctgaagaagtctatcgatatacaactatg 237
Db 181 aatctggaatctcccaaaatcgaaagttaatccctgaagaagtctatcgatatacaactatg 240
QY 238 taagaaactctccacactctctctgacatccgatccgacagaatacattccatccatcc 297
Db 241 taagaaactctccacactctctctgacatccgatccgacagaatacattccatccatcc 300
QY 298 ctgaacaatgaataacacacatccatccatccatccatccatccatccatccatccatcc 357
Db 301 ctgaacaatgaataacacacatccatccatccatccatccatccatccatccatccatcc 360
QY 358 ctgaacaatgaataacacatccatccatccatccatccatccatccatccatccatccatcc 417
Db 361 ctgaacaatgaataacacatccatccatccatccatccatccatccatccatccatccatcc 420
QY 418 gtaaccaatctctcagatgatcaacatctctgactatcaatccgctgactctgctt 477
Db 421 gtaaccaatctctcagatgatcaacatctctgactatcaatccgctgactctgctt 480
QY 478 accatcaccaaatctctcagatgatcaacatctctgactatcaatccgctgactctgctt 537
Db 481 accatcaccaaatctctcagatgatcaacatctctgactatcaatccgctgactctgctt 540
QY 538 cagaacacgataccatccatccatccatccatccatccatccatccatccatccatccatcc 597
Db 541 cagaacacgataccatccatccatccatccatccatccatccatccatccatccatccatcc 600
QY 598 gacggtgtctgacatccatccatccatccatccatccatccatccatccatccatccatcc 657
Db 601 gacggtgtctgacatccatccatccatccatccatccatccatccatccatccatccatcc 660
QY 658 gaactgacgaagaaagaatacgaagactctgacgaacacgaagcgaatctgtatctgt 717
Db 661 gaactgacgaagaaagaatacgaagactctgacgaacacgaagcgaatctgtatctgt 720
QY 718 aaagactctgggtgtgactacccgacgtacgaacacgaagcgaatctgtatctgtac 777
Db 721 aaagactctgggtgtgactacccgacgtacgaacacgaagcgaatctgtatctgtac 780
QY 778 gacgcgaacaaatcgttgcgtacgaacacgaatctgtatctgtatctgtatctgtatctgt 837
Db 781 gacgcgaacaaatcgttgcgtacgaacacgaatctgtatctgtatctgtatctgtatctgt 840
QY 838 ggtccggtgtctctgtatctgtatctgtatctgtatctgtatctgtatctgtatctgtatct 897
Db 841 ggtccggtgtctctgtatctgtatctgtatctgtatctgtatctgtatctgtatctgtatct 900
QY 898 accaaatctatctgaagaatacgaatctgtatctgtatctgtatctgtatctgtatctgtatct 957
Db 901 accaaatctatctgaagaatacgaatctgtatctgtatctgtatctgtatctgtatctgtatct 960
QY 958 gacgtgtatcatcaatctgtatctgtatctgtatctgtatctgtatctgtatctgtatctgtatct 1017
Db 961 gacgtgtatcatcaatctgtatctgtatctgtatctgtatctgtatctgtatctgtatctgtatct 1020
QY 1018 tctcaggtctgtgtgaagaagaatctgtatctgtatctgtatctgtatctgtatctgtatctgt 1077

|||||
Db 1021 tctcagctggttagaagaatctgtctgtcgtgaatcccgagctgtgtaactg 1080
Qy 1078 tctcaggaagttagaataatccaagacagcaggtatccactaacaatgtcaatg 1137
Db 1081 tctcaggaagttagaataatccaagacagcaggtatccactaacaatgtcaatg 1140
Qy 1138 aatctgcagcaacaacagtaacgatacgtgttcacgtgttcaccagttccaacat 1197
Db 1141 aatctgcagcaacaacagtaacgatacgtgttcacgtgttcaccagttccaacat 1200
Qy 1198 atcgtcaactcgtgtgttcccaactcgtgtacaatcgtcagatcgaacgttccctcgcact 1257
Db 1201 atcgtcaactcgtgtgttcccaactcgtgtacaatcgtcagatcgaacgttccctcgcact 1260
Qy 1258 ctgggtgtgtcttgaggatctatcccggtgtatgacggttggtggaacgttcgctgtaa 1317
Db 1261 ctgggtgtgtcttgaggatctatcccggtgtatgacggttggtggaacgttcgctgtaa 1320
Qy 1318 gaattc 1323
Db 1321 gaattc 1326

RESULT 4
AAT29245
ID AAT29245 standard; DNA; 1330 BP.

AC AAT29245;

DT 07-JUL-1996 (first entry)

XX Type A neurotoxin C fragment synthetic gene.

KW Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;
KM Clostridium botulinum; ds.

OS Synthetic.

XX Location/Qualifiers
FT CDS 1..1317
FT 1/*tag= a

PN WO9612802-A1.

XX 02-MAY-1996.

PF 23-OCT-1995; 95WO-US13737.

PR 07-JUN-1995; 95US-0480604.

PR 24-OCT-1994; 94US-0329154.

PR 16-MAR-1995; 95US-0405496.

PR 14-APR-1995; 95US-0422711.

XX (OPHI-) OPHIDIAN PHARM INC.

PI Firca JR, Rink JA, Padhye NV, Stafford DC, Thalley BS;

XX Williams JA;

DR WPI: 1996-230603/23.

XX P-PSDB; AAR95008.

PT Fusion proteins comprising non-toxin protein and part of toxin

PT C. difficile type toxins, and to treat C. difficile intoxication,

PT patric. diarrhoea

XX Example 22: Page 336-38; 434p; English.

XX A synthetic gene (AAT29245) codes for the heavy chain C fragment

CC (AAR95008) of Clostridium botulinum type A neurotoxin (see also

CC AAR95010). Codon usage allowing efficient gene expression in Escherichia

CC coli was utilised. The gene in vector pAlterBot was used to make

CC expression constructs in which fragments of C. difficile toxin A
CC repeat domains were expressed as genetic fusions with the C.
CC botulin C fragment and expressed in E. coli.
XX Sequence 1330 BP; 400 A; 339 C; 246 G; 345 T; 0 other;

Query Match 98.6%; Score 1304.4; DB 17; Length 1330;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 12 gatgttacccttcaactgaatacatcaagaacatcatcaactccatctgaactcgtg 71
Db 12 gctgttacccttcaactgaatacatcaagaacatcatcaactccatctgaactcgtg 71
Qy 72 ctacgaatccaatccactcgtatcgtacctgtctcgtacgcttccaaatcaacatcgtgtc 131
Db 72 ctacgaatccaatccactcgtatcgtacctgtctcgtacgcttccaaatcaacatcgtgtc 131
Qy 132 taaagttaactcgaatccgatacgaagaatccagctgttcaatcgtgaatcttc 191
Db 132 taaagttaactcgaatccgatacgaagaatccagctgttcaatcgtgaatcttc 191
Qy 192 caaaatcgaagttaactcgaagaatgtaatcgtatatacaactctatgtacgaanaactctc 251
Db 192 caaaatcgaagttaactcgaagaatgtaatcgtatatacaactctatgtacgaanaactctc 251
Qy 252 caactcctcttgatccgatacctcgaanaatccaactccatctctcgtgaacatgata 311
Db 252 caactcctcttgatccgatacctcgaanaatccaactccatctctcgtgaacatgata 311
Qy 312 caactcctcttgatccgatacctcgaanaatccaactccatctctcgtgaacatgata 371
Db 312 caactcctcttgatccgatacctcgaanaatccaactccatctctcgtgaacatgata 371
Qy 372 aatcatctgactcgtcgaagacactcgaagaatccaacacgctgtgtatccaatactc 431
Db 372 aatcatctgactcgtcgaagacactcgaagaatccaacacgctgtgtatccaatactc 431
Qy 432 tgaagtatcaacatctcgtatcatcaatctcgtggaatctcgttaccatccaaca 491
Db 432 tgaagtatcaacatctcgtatcatcaatctcgtggaatctcgttaccatccaaca 491
Qy 492 tctgtgaataactcgaatccaatctcatcaacgctcgtatcgaacgaacacgactc 551
Db 492 tctgtgaataactcgaatccaatctcatcaacgctcgtatcgaacgaacacgactc 551
Qy 552 caactcgtgtaacatccaactccttaataacatcgttcaaacctggaacgtgttcgtga 611
Db 552 caactcgtgtaacatccaactccttaataacatcgttcaaacctggaacgtgttcgtga 611
Qy 612 caactcgtgtaacatccaactccttaataacatcgttcaaacctggaacgtgttcgtga 671
Db 612 caactcgtgtaacatccaactccttaataacatcgttcaaacctggaacgtgttcgtga 671
Qy 672 agaaatcaagaacgctgaagacacacgaatcctcgtatcctcgaagaactcctcgtg 731
Db 672 agaaatcaagaacgctgaagacacacgaatcctcgtatcctcgaagaactcctcgtg 731
Qy 732 tgactactcgtcagtaacgaacacgtaactacatcgttcaaacctggaacgtgttcgtga 791
Db 732 tgactactcgtcagtaacgaacacgtaactacatcgttcaaacctggaacgtgttcgtga 791
Qy 792 cgttgaacgaacacgtaactacatcgttcaaacctggaacgtgttcgtga 851
Db 792 cgttgaacgaacacgtaactacatcgttcaaacctggaacgtgttcgtga 851
Qy 852 tgttgaactacacacgtaactacatcgttcaaacctggaacgtgttcgtga 911
Db 852 tgttgaactacacacgtaactacatcgttcaaacctggaacgtgttcgtga 911
Qy 912 caagaatacgcgtctgtgaacgaacatcgttccgaacaaatgactcgtgtatcat 971
Db 912 caagaatacgcgtctgtgaacgaacatcgttccgaacaaatgactcgtgtatcat 971

```

Db 912 caagaataacgcgtctgtaacaagaacaataatcgttcgcaacaatgctgtatacat 971
Qy 972 caatgtgtagttaagaacaagaataacacgctctgtctaccatgctctcaggctgtgt 1031
Db 972 caatgtgtagttaagaacaagaataacacgctctgtctaccatgctctcaggctgtgt 1031
Qy 1032 agaaagaatcttctgtctctgtgaatcccgagcgttgtgtatctgtctcaggtaagt 1091
Db 1032 agaaagaatcttctgtctctgtgaatcccgagcgttgtgtatctgtctcaggtaagt 1091
Qy 1092 aatgaataccaagaacgacgaggtatcactaaatgaataatgaatctgcaagacaa 1151
Db 1092 aatgaataccaagaacgacgaggtatcactaaatgaataatgaatctgcaagacaa 1151
Qy 1152 caatgtgtacgataatcgtttcactcgtttccacagttcaacaataatcgttaactgt 1211
Db 1152 caatgtgtacgataatcgtttcactcgtttccacagttcaacaataatcgttaactgt 1211
Qy 1212 tgcgtccaactgtgtacaatcgtcagatcgaaagcttccctcgcacactgtgtgtcttg 1271
Db 1212 tgcgtccaactgtgtacaatcgtcagatcgaaagcttccctcgcacactgtgtgtcttg 1271
Qy 1272 ggaatccatcccggttgatgaacggttggaagctgcgctgaa 1317
Db 1272 ggaatccatcccggttgatgaacggttggaagctgcgctgaa 1317

RESULT 5
AAV30571
ID AAV30571 standard; DNA; 1330 BP.
XX
AC AAV30571;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C gene in palterBot.
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulinum; ds.
XX
OS Clostridium botulinum serotype A.
XX
FH Key Location/Qualifiers
FT CDS 1..1317
FT misc_difference 1..6 /tag= a
FT /tag= b
FT /note= "palter vector-derived nucleotides
FT (encode Met-Ala)"
XX
PN W09808540-A1.
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US15394.
XX
PR 28-AUG-1996; 96US-0704159.
XX
PA (OPH1-) OPHIDIAN PHARM INC.
XX
PI Thalleley BS, Williams JA.
XX
DR WPI; 1998-230234/20.
DR P-PSDB; AAW68389.
XX
PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with Clostridial toxin
XX
PS Example 22; Page 262-263; 428pp; English.
XX
CC This is the DNA sequence of the Clostridium botulinum serotype A
CC toxin C-fragment gene contained in plasmid palterBot. Recombinant

```

```

CC C-fragment proteins have been produced in Escherichia coli as
CC fusion proteins with either maltose binding protein or
CC Clostridium difficile type A toxin (see AAW68387). The invention
CC relates to recombinant proteins derived from C. botulinum toxins.
CC Methods are provided which allow for the isolation of soluble
CC recombinant proteins free of significant endotoxin contamination.
CC Preferred hosts for production of recombinant toxin proteins are
CC insect cells and yeast cells. The recombinant toxin proteins are
CC used as immunogens for the production of vaccines and antitoxins
CC that are useful in the treatment of humans and animals at risk of
CC intoxication with clostridial toxin.
XX
SQ Sequence 1330 BP; 400 A; 339 C; 246 G; 345 T; 0 other:

Query Match 98.6%; Score 1304.4; DB 19; Length 1330;
Best local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 12 gatgtctaccttaactgaatatacaagaacatcatcaatcctcaatcctgaacctg 71
Db 12 gctgtctaccttaactgaatatacaagaacatcatcaatcctcaatcctgaacctg 71
Qy 72 ctacgaatccaatcacctgtagtgcacctgtctcgtctacgcttccaaaatcaaatcggttc 131
Db 72 ctacgaatccaatcacctgtagtgcacctgtctcgtctacgcttccaaaatcaaatcggttc 131
Qy 132 taaagttaacttcgatccgacacgacgaagaatcagatccagctgttcaatctggaatcttc 191
Db 132 taaagttaacttcgatccgacacgacgaagaatcagatccagctgttcaatctggaatcttc 191
Qy 192 caaatcgaagttatccctgaagaatgctatcgatatacaactatgtagaanaactcttc 251
Db 192 caaatcgaagttatccctgaagaatgctatcgatatacaactatgtagaanaactcttc 251
Qy 252 cactctctctgtgatacgcgtatccgaataatcctcaatcctcctcgtgaacaatgaata 311
Db 252 cactctctctgtgatacgcgtatccgaataatcctcaatcctcctcgtgaacaatgaata 311
Qy 312 caccatcatcaatctgatactgaagaacaattctgtgtgaaagtatactctgatacgaatgta 371
Db 312 caccatcatcaatctgatactgaagaacaattctgtgtgaaagtatactctgatacgaatgta 371
Qy 372 aatcatctggaactctgaggaacacacgaagaatacaacagcgtgtgtatccaatc 431
Db 372 aatcatctggaactctgaggaacacacgaagaatacaacagcgtgtgtatccaatc 431
Qy 432 tcagatgatacaatctctgactacataatcgctggtatcttggttacacctcaacaaga 491
Db 432 tcagatgatacaatctctgactacataatcgctggtatcttggttacacctcaacaaga 491
Qy 492 tcgtctgaataactccaanaatctacatacaagcgctctgatcgcagacaagaacgatactc 551
Db 492 tcgtctgaataactccaanaatctacatacaagcgctctgatcgcagacaagaacgatactc 551
Qy 552 caatctgggttaacatccacgctcttaataacatcatggttcaactgtgaaggtgtgtgta 611
Db 552 caatctgggttaacatccacgctcttaataacatcatggttcaactgtgaaggtgtgtgta 611
Qy 612 cactcacgcgtatcatctggtatcaatactcaatctgttgcgaagaagaatctgaacgaaa 671
Db 612 cactcacgcgtatcatctggtatcaatactcaatctgttgcgaagaagaatctgaacgaaa 671
Qy 672 agaatcaagaacgtctgacacacacagttccaattctgtatctctgaaagactctgggg 731
Db 672 agaatcaagaacgtctgacacacacagttccaattctgtatctctgaaagactctgggg 731
Qy 732 tgactactgtagtacgacaacacgtaactatcatgctgaatctgttaacgatccgaacaata 791
Db 732 tgactactgtagtacgacaacacgtaactatcatgctgaatctgttaacgatccgaacaata 791
Qy 792 cgttgagcttaacaagttagtatactcggtgtatcatgtactctgaaaggtccggtgttc 851
Db 792 cgttgagcttaacaagttagtatactcggtgtatcatgtactctgaaaggtccggtgttc 851

```

Dh 792 cgtgacgtcaacaatgtagtgcgcggttacatgtaacctgaaggctcgcgtgttc 851
Qy 852 tgtttgctactacaactactactactactcttccctgtacggtgttaccatataat 911
Dh 852 tgtttgctactacaactactactactactcttccctgtacggtgttaccatataat 911
Qy 912 caagaatacgcgtctgttgaagaacataatgcttcgaacaatgtatcgtataat 971
Dh 912 caagaatacgcgtctgttgaagaacataatgcttcgaacaatgtatcgtataat 971
Qy 972 caatgtgttagtgaagaacataatgcttcgaacaatgtatcgtataat 1031
Dh 972 caatgtgttagtgaagaacataatgcttcgaacaatgtatcgtataat 1031
Qy 1032 agaaaagcttctgtctgtcgtgaataatccgcgagcttgtaactcgtctcagttgt 1091
Dh 1032 agaaaagcttctgtctgtcgtgaataatccgcgagcttgtaactcgtctcagttgt 1091
Qy 1092 aatgaatacgaagaacgcagcaggtatcacatacaaatgaataatcgcagagaa 1151
Dh 1092 aatgaatacgaagaacgcagcaggtatcacatacaaatgaataatcgcagagaa 1151
Qy 1152 caatgtgaatgataatcgttcaatcgttccacagttcaacaataatcgttaactgt 1211
Dh 1152 caatgtgaatgataatcgttcaatcgttccacagttcaacaataatcgttaactgt 1211
Qy 1212 tgcctcaactcgtcgaatcgtcgaatcgtcgaatcgtcgaatcgtcgaatcgt 1271
Dh 1212 tgcctcaactcgtcgaatcgtcgaatcgtcgaatcgtcgaatcgtcgaatcgt 1271
Qy 1272 ggaagtcacccgcgttgatgaacgttgggtgaacgtccgctgttaa 1317
Dh 1272 ggaagtcacccgcgttgatgaacgttgggtgaacgtccgctgttaa 1317

RESULT 6
AAZ87212
ID AAZ87212 standard; DNA; 1338 BP.
AC AAZ87212;
XX
DT 08-MAY-2000 (first entry)
XX
DE DNA encoding synthetic BONT serotype A (BONTA) Hc fragment.
XX
KW Botulinum neurotoxin; heavy chain; BONT; serotype A;
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening; ds.
XX
OS Clostridium botulinum.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 9..1325
FT /tag= "A
FT /product= "Synthetic botulinum neurotoxin serotype A
FT (BONTA) heavy chain C-terminal fragment (Hc)"
XX
XX WO200002524-A2.
XX
XX 20-JAN-2000.
XX
XX 09-JUL-1999; 99WO-US15570.
XX
XX 10-JUL-1998; 98US-0092416.
XX 12-MAY-1999; 99US-0133870.
XX
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
XX Lee JS, Pushko P, Smith JF, Parker M, Dertbaugh MT, Smith L;
XX WPI; 2000-160827/14.
XX
XX P-PSDB; AAY77134.
DR

XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum -
PS
XX Disclosure: Page 54; 54pp; English.
XX
CC The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BONT) serotypes A, B, C, D, E, F
CC and G (BONTA-BONTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a vector, and at least one nucleic acid
CC fragment comprising a C-terminal heavy chain fragment (Hc) from BONT
CC serotypes A-G. In preferred embodiments of the invention, the vector is
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC this vector results in the production of large amounts of a protein
CC encoded by a sequence cloned into the replicon. The constructs are used
CC to produce vaccines against botulism. The proteins can also be used as
CC diagnostic tools for the diagnosis of botulism. The transformed host
CC cells can be used to analyse the effectiveness of drugs and agents which
CC inhibit toxin effects. The vaccine currently used against botulism is
CC dangerous and expensive to produce, and contains formalin, which is very
CC painful for the recipient. Also, the vaccine is incomplete, in that only
CC 5 of the 7 serotypes are represented in the formulation. The novel
CC vaccine of overcomes these problems, as it is easily purified, and
CC available in large quantities. It is also expressed in the lymph nodes
CC for a better immune response. Sequences AAZ87212-287217 represent
CC synthetic DNA sequences encoding BONT Hc fragments used in the present
CC invention. These were optimised for codon usage for expression in yeast.
XX
SQ Sequence 1338 BP; 401 A; 342 C; 249 G; 346 T; 0 other;

Query Match 98.6%; Score 1304.4; DB 21; Length 1338;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 12 gatgtctacctcaactgaatatacaagaacatcaatcaatccatccatcgtgaactcgt 71
Dh 20 gctgttactctcactgaatatacaagaacatcaatcaatccatccatcgtgaactcgt 79
Qy 72 ctagaatccaatcacctgatacgaacctgtctcgtcgtcgtccccaatacaatcgtgttc 131
Dh 80 ctacgaatccaatcacctgatacgaacctgtctcgtcgtcgtccccaatacaatcgtgttc 139
Qy 132 taaagtaacttcgatcgcgatcgacaagaataagatccgcgtgttaactcgtgaacttc 191
Dh 140 taaagtaacttcgatcgcgatcgacaagaataagatccgcgtgttaactcgtgaacttc 199
Qy 192 caaatcgaagtatctctaagaatgctatcgatacaactctatgtaagaactcttc 251
Dh 200 caaatcgaagtatctctaagaatgctatcgatacaactctatgtaagaactcttc 259
Qy 252 caactcctcttgatcgcgtatccgcgaataactcaactccatctctctgaacaatgata 311
Dh 260 caactcctcttgatcgcgtatccgcgaataactcaactccatctctctgaacaatgata 319
Qy 312 caccatcgaactgataatgaagaataatcgtgtggaagaatctctgaactcgtgta 371
Dh 320 caccatcgaactgataatgaagaataatcgtgtggaagaatctctgaactcgtgta 379
Qy 372 aatcatctgactctgagagacactcagaagaatacaacagcgtgtgtatctcaatactc 431
Dh 380 aatcatctgactctgagagacactcagaagaatacaacagcgtgtgtatctcaatactc 439
Qy 432 tcagatgatacaatctctgactatacaatcgtcgtgatacttcgttaacatacaaca 491
Dh 440 tcagatgatacaatctctgactatacaatcgtcgtgatacttcgttaacatacaaca 499
Qy 492 tcgtcgaataactccaataatctacatacaacgcgcgtcgtatgaccagaacaacgcttc 551
Dh 500 tcgtcgaataactccaataatctacatacaacgcgcgtcgtatgaccagaacaacgcttc 559
Qy 552 caatcgtgtaacatccagcttctaataacatcagttcaaacgtgacggtgtgtcgtga 611

```

Db      560 caatcgggtaaacatccacgcttctaataacatcgttaacatcgtgacgtgtgctgtga
Qy      612 caactaccgctacatctgtgtcaataactcaatctgttgcagaagaactgacgaa
Db      620 cactaccgctacatctgtgtcaataactcaatctgttgcagaagaactgacgaa
Qy      672 agaatacaaaagacctgacgacaacacagttccaaattctgtatccctgaagaactctggg
Db      680 agaatacaaaagacctgacgacaacacagttccaaattctgtatccctgaagaactctggg
Qy      732 tgactaccgctgacgacaacacgcttactacgctgaactctgtatccctgaagaactctggg
Db      740 tgactaccgctgacgacaacacgcttactacgctgaactctgtatccctgaagaactctggg
Qy      792 cgttgcgtcaacaatgtagtataccgctgtacatgtagtactaaaggtccgctgtgttc
Db      800 cgttgcgtcaacaatgtagtataccgctgtacatgtagtactaaaggtccgctgtgttc
Qy      852 tgtatgactaccacatctactcgtgaactcttccctgtacacgtgtgtacaaattcatcat
Db      860 tgtatgactaccacatctactcgtgaactcttccctgtacacgtgtgtacaaattcatcat
Qy      912 caagaataacgctgtgtgtacagaacaatctcgttgcagaacaatgtagtctgtatcat
Db      920 caagaataacgctgtgtgtacagaacaatctcgttgcagaacaatgtagtctgtatcat
Qy      972 caatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Db      980 caatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Qy      1032 agaaaagatctgtctgtctgtgaatcccgagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Db      1040 agaaaagatctgtctgtctgtgaatcccgagcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Qy      1092 aatgaataatccaaagacaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Db      1100 aatgaataatccaaagacaggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Qy      1152 caatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Db      1160 caatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Qy      1212 tgccttcaactgtgtacatctgtcatgtgaacgttccctgcgaactctgtgtgtgtgtgt
Db      1220 tgccttcaactgtgtacatctgtcatgtgaacgttccctgcgaactctgtgtgtgtgt
Qy      1272 ggaatcattcccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt
Db      1280 ggaatcattcccggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt

```

RESULT 7
AAV30576
ID AAV30576 standard; DNM; 1351 BP.
AC AAV30576;
DF 07-DEC-1998 (first entry)
XX
XX
DE Clostridium botulinum toxin A fragment C gene in pHisBotA(syn).
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
XX botulinism; ds.
XX
OS Clostridium botulinum serotype A.
XX Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 1..1338
FT /tag= a
FT
PN WO9808540-A1.

```

XX      05-MAR-1998.
PD      28-AUG-1997; 97MO-US15394.
PF      28-AUG-1996; 96US-0704159.
XX      (OPHI-) OPHIDIAN PHARM INC.
XX      Thalley BS, Williams JA.
XX      WPI: 1998-230234/20.
DR      P-PSDB; AAM68391.
PT      Host cell, containing recombinant expression vector encoding
XX      Clostridium botulinum type B or E toxin - useful to treat humans
XX      and other animals at risk of intoxication with clostridial toxin.
XX      Example 29; Page 279-281; 428pp; English.
XX      This is the DNA sequence of the Clostridium botulinum serotype A
XX      toxin C fragment gene contained in plasmid pHisBotA(syn). The
XX      encoded toxin A polypeptide (see AAM68391) has a histidine-tagged
XX      N-terminal extension. The vector was used to express native
XX      (i.e. non-fusion) soluble C fragment in Escherichia coli host
XX      cells. The invention relates to recombinant proteins derived from
XX      C. botulinum toxins. Methods are provided which allow for the
XX      isolation of soluble recombinant proteins free of significant
XX      endotoxin contamination. Preferred hosts for production of
XX      recombinant proteins are E. coli, insect cells and yeast cells.
XX      The recombinant toxins are used as immunogens for the production
XX      of vaccines and antitoxins that are useful in the treatment of
XX      humans and animals at risk of intoxication with clostridial toxin.
XX      Sequence 1351 BP; 407 A; 348 C; 247 G; 349 T; 0 other:

```

Query Match 98.6%; Score 1304.4; DB 19; Length 1351;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      12 gatgtctacttactgaatatacagaagaacatcatcaatcattcattcgtgacgtg 71
Db      33 gctgttcaacttactgaatatacagaagaacatcatcaatcattcattcgtgacgtg 92
Qy      72 ctacgaatccaatcaactgtatcgacctgtctgtctgtgtgtgtgtgtgtgtgtgtgt 131
Db      93 ctacgaatccaatcaactgtatcgacctgtctgtgtgtgtgtgtgtgtgtgtgtgtgt 152
Qy      132 taaagttaacttcgatccgacgacgaagaatcagatccagctgttcaatctggaatttc 212
Db      153 taaagttaacttcgatccgacgacgaagaatcagatccagctgttcaatctggaatttc 212
Qy      192 caaaatcgaagtattccctggaagatgtatcgtatatacaactcatgtacgaaactcttc 251
Db      213 caaaatcgaagtattccctggaagatgtatcgtatatacaactcatgtacgaaactcttc 272
Qy      252 cactctcttctgtgtatccgaaatactcaactcaatctctgtgtgtgtgtgtgtgtgtgt 311
Db      273 cactctcttctgtgtatccgaaatactcaactcaatctctgtgtgtgtgtgtgtgtgtgt 332
Qy      312 caccatcatcaactgtatgaaacaaatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 371
Db      333 caccatcatcaactgtatgaaacaaatctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 392
Qy      372 aatcatctgtgactgtgaggaacactcaggaatacaaacacgtgtgtgtgtgtgtgtgtgt 431
Db      393 aatcatctgtgactgtgaggaacactcaggaatacaaacacgtgtgtgtgtgtgtgtgtgt 452
Qy      432 tcagatgatcaacaatctctactacatcaatctcgtgtatcttgcgttaccatccacaaca 491
Db      453 tcagatgatcaacaatctctactacatcaatctcgtgtatcttgcgttaccatccacaaca 512

```

QY 492 tgcgtgataataatccaaatctacatcaacgcgctctgatacgacaagaacgatctc 551
|||||
Db 513 tgcgtgataataatccaaatctacatcaacgcgctctgatacgacaagaacgatctc 572
QY 552 caatctgggttaacatccacgctcttaataacatcgttcaaacctgagcggttgcgtga 611
|||||
Db 573 caatctgggttaacatccacgctcttaataacatcgttcaaacctgagcggttgcgtga 632
QY 612 caatcaacgcttaacatcgttcaaacctgagcggttgcgtga 671
|||||
Db 633 caatcaacgcttaacatcgttcaaacctgagcggttgcgtga 692
QY 672 agaaatcaacgcttaacatcgttcaaacctgagcggttgcgtga 731
|||||
Db 693 agaaatcaacgcttaacatcgttcaaacctgagcggttgcgtga 752
QY 732 tgaactactgacgtacgacaaacgcttaacatcgttcaaacctgagcggttgcgtga 791
|||||
Db 753 tgaactactgacgtacgacaaacgcttaacatcgttcaaacctgagcggttgcgtga 812
QY 792 cgttgacgtcaacaaatgaggtatccggttgcgttaacatcgttcaaacctgagcggttgc 851
|||||
Db 813 cgttgacgtcaacaaatgaggtatccggttgcgttaacatcgttcaaacctgagcggttgc 872
QY 852 tgtatgactacaaacatctacatcgttcaaacctgagcggttgcgttaacatcgttcaaac 911
873 tgtatgactacaaacatctacatcgttcaaacctgagcggttgcgttaacatcgttcaaac 932
QY 912 caagaataacgcttgcgttaacaaagacaaatcgttgcgaacaatgagcggttgcgttaacat 971
Db 933 caagaataacgcttgcgttaacaaagacaaatcgttgcgaacaatgagcggttgcgttaacat 992
QY 972 caatcttgatgtaagaacaaagacaaatcgttgcgaacaatgagcggttgcgttgcgttgc 1031
Db 993 caatcttgatgtaagaacaaagacaaatcgttgcgaacaatgagcggttgcgttgcgttgc 1052
QY 1032 agaaagaatctgtctgcttggaatacccgagcggttgcgttaacatcgttcaaacctgag 1091
Db 1053 agaaagaatctgtctgcttggaatacccgagcggttgcgttaacatcgttcaaacctgag 1112
QY 1092 aatgaataccaaagacaaagcggttgcgttaacaaatggaatacgttcgaagacaa 1151
Db 1113 aatgaataccaaagacaaagcggttgcgttaacaaatggaatacgttcgaagacaa 1172
QY 1152 caatgtaacgatacgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc 1211
Db 1173 caatgtaacgatacgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc 1232
QY 1212 tgcgtccaaactgtaacaaatcgttcagatcgaaacgttgcgttgcgttgcgttgcgttgc 1271
Db 1233 tgcgtccaaactgtaacaaatcgttcagatcgaaacgttgcgttgcgttgcgttgcgttgc 1292
QY 1272 ggaagtcacatccggttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc 1317
Db 1293 ggaagtcacatccggttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgcgttgc 1338

RESULT 8

AAT29246
ID AAT29246 standard; DNA: 1402 BP.

AC AAT29246;

XX 07-JUL-1996 (first entry)

DE Type A neurotoxin C fragment-polyhistidine tag gene fusion.

XX Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;

KW Clostridium botulinum; polyhistidine; vector; pETHsa; pHisBot; ds.

OS Synthetic.

XX Key Location/Qualifiers

FH

FT CDS 1..1317
FT /tag= a
FT /product= pHisBot fusion protein
PN W09612802-AL.
PD 02-MAY-1996.
XX 23-OCT-1995; 95MO-US13737.
XX 07-JUN-1995; 95US-0480604.
PR 24-OCT-1994; 94US-0329154.
PR 16-MAR-1995; 95US-0405496.
PR 14-APR-1995; 95US-0422711.
XX (OPHI-) OPHIDIAN PHARM INC.
XX Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;
PI Williams JA;
XX MPI: 1996-230603/23.
DR P-PSDB; AAR95009.
XX Fusion proteins comprising non-toxin protein and part of toxin
PT useful to form anti-toxins against Clostridium botulinum type A, and
PT C. difficile type toxins, and to treat C. difficile intoxication,
PT partic. diarrhoea
XX Example 24; Page 340-342; 434pp; English.
PS A nucleotide sequence (AAT29246) present in vector pETHsa encodes
CC the pHisBot fusion protein (AAR95009) comprising a polyhistidine
CC affinity tag and fragment C (see also AAR95008) of the Clostridium
CC botulinum type A neurotoxin. The pHisBot protein was expressed
CC in Escherichia coli as a soluble protein and was purified by
CC metal chelate affinity chromatography to obtain a product free
CC of endotoxin contamination that may be useful as an immunogen
CC in vaccine compns.
SQ Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other;

Query Match 98.6%; Score 1304.4; DB 17; Length 1402;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 gatgttacctcaactgataatcaagaacatcatcatcattccattcaactgcg 71
Db 84 gctgttaccttcaactgataatcaagaacatcatcatcattccattcaactgcg 143
QY 72 ctagaatacaatcaactgataatcaagaacatcatcatcattccattcaactgcg 131
Db 144 ctagaatacaatcaactgataatcaagaacatcatcatcattccattcaactgcg 203
QY 132 taaagttaacttcgatacgcgacgacgaagaatcgaatcgaatcgaatcgaatcgaatcga 191
Db 204 taaagttaacttcgatacgcgacgacgaagaatcgaatcgaatcgaatcgaatcgaatcga 263
QY 192 caaatcgaagttaacttcgaagaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 251
Db 264 caaatcgaagttaacttcgaagaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 323
QY 252 caactcctctgatacgcgatacgcgatacgcgatacgcgatacgcgatacgcgatacgcgata 311
Db 324 caactcctctgatacgcgatacgcgatacgcgatacgcgatacgcgatacgcgatacgcgata 383
QY 312 caacatcatcaactgatacgcgatacgcgatacgcgatacgcgatacgcgatacgcgatacgcg 371
Db 384 caacatcatcaactgatacgcgatacgcgatacgcgatacgcgatacgcgatacgcgatacgcg 443
QY 372 aatcatctgactctgcaggaacatcgaagaatcaacaacgctgtgtatccaatactc 431
Db 444 aatcatctgactctgcaggaacatcgaagaatcaacaacgctgtgtatccaatactc 503

QY 432 tcaagatgatacaacatctctgactacatacaatcgctgagatcttcgttaccatccacaaca 491
 |||||
 DB 504 tcagatgatacaacatctctgactacatacaatcgctgagatcttcgttaccatccacaaca 563
 QY 492 tggctggaataactccaataatctacatacaacggtcgctgagatcttcgttaccatccacaaca 551
 |||||
 DB 564 tcgctggaataactccaataatctacatacaacggtcgctgagatcttcgttaccatccacaaca 623
 QY 552 caatctggttaacatcaacggtcttataataacatcgttcaaatctggaacggtctgtctgta 611
 |||||
 DB 624 caatctggttaacatcaacggtcttataataacatcgttcaaatctggaacggtctgtctgta 683
 QY 612 cactcacgctacatactgtaatacaataactcaatctgttcgcgaagaagactggaagaana 671
 |||||
 DB 684 cactcacgctacatactgtaatacaataactcaatctgttcgcgaagaagactggaagaana 743
 QY 672 agaaatcaagaagactgtaagacacacggtccaaattctgtatctctggaagactctgggg 721
 |||||
 DB 744 agaaatcaagaagactgtaagacacacggtccaaattctgtatctctggaagactctgggg 803
 QY 732 tgactacctgcaagtaagacacacggtctacatactgctggaatctgtacgatactcgacaata 791
 |||||
 DB 804 tgactacctgcaagtaagacacacggtctacatactgctggaatctgtacgatactcgacaata 863
 QY 792 cgttgacgttcaacaatgtagtaaccggttatactgtaacgttgaagaagctcggtgttc 851
 |||||
 DB 864 cgttgacgttcaacaatgtagtaaccggttatactgtaacgttgaagaagctcggtgttc 923
 QY 852 tgtatgactacacacatctacatcttccctgtaacgttgcgtgtaacacatctcat 911
 |||||
 DB 924 tgtatgactacacacatctacatcttccctgtaacgttgcgtgtaacacatctcat 983
 QY 912 caagaataacgctgtctgtaacaagaacataatcgttgcgaacatgactggtatatacat 971
 |||||
 DB 984 caagaataacgctgtctgtaacaagaacataatcgttgcgaacatgactggtatatacat 1043
 QY 972 caatgtttagttaaagaacaagaataacggtctggttaccatgcttctgaagctgtgt 1031
 |||||
 DB 1044 caatgtttagttaaagaacaagaataacggtctggttaccatgcttctgaagctgtgt 1103
 QY 1032 agaaagatcttctgctctgtaaatcccgagcgttctgtaactgtctcaggttagttgt 1091
 |||||
 DB 1104 agaaagatcttctgctctgtaaatcccgagcgttctgtaactgtctcaggttagttgt 1163
 QY 1092 aatgaaatccaaagacgaccaggttatacaatacaaatgcaaaatgaaatctgcaagcaaa 1151
 |||||
 DB 1164 aatgaaatccaaagacgaccaggttatacaatacaaatgcaaaatgaaatctgcaagcaaa 1223
 QY 1152 caatgtttagttaaagaacaagaataacggtctggttaccatgcttctgaagctgtgt 1211
 |||||
 DB 1224 caatgtttagttaaagaacaagaataacggtctggttaccatgcttctgaagctgtgt 1283
 QY 1212 tgcctccaactggttacaatcgttacaatcgaaacgttccctcgcacactggtgtgctctg 1271
 |||||
 DB 1284 tgcctccaactggttacaatcgttacaatcgaaacgttccctcgcacactggtgtgctctg 1343
 QY 1272 gggagttacatcccggttgaagaagcgttgggtggaacgttcgcgtgttaa 1317
 |||||
 DB 1344 gggagttacatcccggttgaagaagcgttgggtggaacgttcgcgtgttaa 1389

RESULT 9

AAV30572 standard; DNA; 1402 BP.

AAV30572;

07-DEC-1998 (first entry)

Clostridium botulinum toxin A fragment C gene in phisBot.

AntiToxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;

KW botulinm: ds.
 OS Clostridium botulinum serotype A.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1389
 FT /tag= a
 XX
 PN W09808540-A1.
 XX
 PD 05-MAR-1998.
 XX
 PF 28-AUG-1997; 97WO-US15394.
 XX
 PR 28-AUG-1996; 96US-0704159.
 XX
 PA (OPHI-) OPHIDIAN PHARM INC.
 XX
 PI Thalley BS, Williams JA.
 DR
 DR MPI: 1998-230234/20.
 PI P-PSDB; AAM68390.
 XX
 PT Host cell containing recombinant expression vector encoding
 PT Clostridium botulinum type B or E toxin - useful to treat humans
 PT and other animals at risk of intoxication with clostridial toxin
 PS Example 24; Page 265-267; 428pp; English.
 XX
 CC This is the DNA sequence of the Clostridium botulinum serotype A
 CC toxin C fragment gene contained in plasmid phisBot. The encoded
 CC toxin A polypeptide (see AAM68390) has a histidine-tagged
 CC N-terminal extension. The vector was used to express native
 CC (i.e. non-fusion) soluble C fragment in Escherichia coli host
 CC cells. The invention relates to recombinant proteins derived from
 CC Clostridium toxins. Methods are provided which allow for the
 CC isolation of soluble recombinant proteins free of significant
 CC endotoxin contamination. Preferred hosts for production of
 CC recombinant proteins are E. coli, insect cells and yeast cells.
 CC The recombinant toxins are used as immunogens for the production
 CC of vaccines and antitoxins that are useful in the treatment of
 CC humans and animals at risk of intoxication with clostridial toxin.
 XX
 SQ Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other;

Query Match 98.6%; Score 1304.4; DB 19; Length 1402;
 Best local similarity 99.9%; Pred. No. 0;
 Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 gatgtacaccttaactgaataacatacaagaacatcatcaatccatccatcgactgctg 71
 |||||
 DB 84 gctgtacaccttaactgaataacatacaagaacatcatcaatccatccatcgactgctg 143
 QY 72 ctacgaatccaactcactggttgaacctgttctgctctacgcttccaataatcaaatcgcttc 131
 |||||
 DB 144 ctacgaatccaactcactggttgaacctgttctgctctacgcttccaataatcaaatcgcttc 203
 QY 132 taaagttaactctgatacgcatacagaagaatcgataccagctgttcaatctggaatcttc 191
 |||||
 DB 204 taaagttaactctgatacgcatacagaagaatcgataccagctgttcaatctggaatcttc 263
 QY 192 caaatcgaaagttaacatccatgaagaatgctatcgtatatacaactatgtaagaactcttc 251
 |||||
 DB 264 caaatcgaaagttaacatccatgaagaatgctatcgtatatacaactatgtaagaactcttc 323
 QY 252 caactcttctgatacgcatacagaagaatcactcaactcaatctctctggaataatgaata 311
 |||||
 DB 324 caactcttctgatacgcatacagaagaatcactcaactcaatctctctggaataatgaata 383
 QY 312 caccatcatcaactgatacgaagaacatctctgttgaagaatgatactctggaactggtga 371
 |||||


```

OY 192 caaaatcgaaagtatccctgaagaatgctatcgatacaactctatgtaacgaanaactctc 251
    |||||
DB 200 caaaatcgaaagtatccctgaagaatgctatcgatacaactctatgtaacgaanaactctc 259
OY 252 caactctcttgatctcgatccgaataactcaactcatctctctgaacaaatgaata 311
    |||||
DB 260 caactctcttgatctcgatccgaataactcaactcatctctctgaacaaatgaata 318
OY 312 caacatcatcaacttgatctgaagaacaaatctggtctggaagaatctctctgaacacgtga 371
    |||||
DB 319 caacatcatcaacttgatctgaagaacaaatctggtctggaagaatctctctgaacacgtga 378
OY 372 aatcatctggactctgacgaagacatcgaagaatcaaacagcgtgtgtatcaatactc 431
    |||||
DB 379 aatcatctggactctgacgaagacatcgaagaatcaaacagcgtgtgtatcaatactc 438
OY 432 tcagatgatcaaatctctgactacataatgctgtgattcttggttaccctccacaaca 491
    |||||
DB 439 tcagatgatcaaatctctgactacataatgctgtgattcttggttaccctccacaaca 498
OY 492 tcgtctgaataactccaaaaatctacatcaacggtctgacacgaagaacgcatctc 551
    |||||
DB 499 tcgtctgaataactccaaaaatctacatcaac-gcgtctgattgacacgaagaacgcatctc 557
OY 552 caactcgggtlaaacatccacgctcttaataacatcatgttcaaacctggacggtgtcgtga 611
    |||||
DB 558 caactcgggtlaaacatccacgctcttaataacatcatgttcaaacctggacggtgtcgtga 616
OY 612 caactcgggtlaaacatccacgctcttaataacatcatgttcaaacgaaagtgaagaanaa 671
    |||||
DB 617 caactcgggtlaaacatccacgctcttaataacatcatgttcaaacgaaagtgaagaanaa 676
OY 672 agaaatcaaaagacctgtgacgacaaacagttccaaattctgtatccctgaaagacctcgggg 731
    |||||
DB 677 agaaatcaaaagacctgtgacgacaaacagttccaaattctgtatccctgaaagacctcgggg 736
OY 732 tgactcctcggcgaatgaacaaacgctactacatgctgaatctgttaagatcccgacaata 791
    |||||
DB 737 tgactcctcggcgaatgaacaaacgctactacatgctgaatctgttaagatcccgacaata 796
OY 792 cgttcgctgaacaaatgtaagatctcgcggttatacgttaccctgaagaagtcgcggttc 851
    |||||
DB 797 cgttcgctgaacaaatgtaagatctcgcggttatacgttaccctgaagaagtcgcggttc 856
OY 852 tgtatgatactcaacaatctacatctcctcctgtacccgtgtgtaacaaatctcat 911
    |||||
DB 857 tgtatgatactcaacaatctacatctcctcctgtacccgtgtgtaacaaatctcat 916
OY 912 caagaataacggtctgtaagaagaacaaatctcgttcgacaagaatgattgtatatacat 971
    |||||
DB 917 caagaataacggtctgtaagaagaacaaatctcgttcgacaagaatgattgtatatacat 973
OY 972 caatgttgaatgaagaacaaagaataacgctcgtgtactacaatgctcttaagcctgtgt 1031
    |||||
DB 974 caatgttgaatgaagaacaaagaataacgctcgtgtactacaatgctcttaagcctgtgt 1033
OY 1032 agaaaagatcttgcctcctggaatcccggaacgttgtgaatctgtctcaagtagtgt 1091
    |||||
DB 1034 agaaaagatcttgcctcctggaatcccggaacgttgtgaatctgtctcaagtagtgt 1093
OY 1092 aatgaataatcaagaacgacgaggtatcactaaacaatgaataatgaatctgcagagcaa 1151
    |||||
DB 1094 aatgaataatcaagaacgacgaggtatcactaaacaatgaataatgaatctgcagagcaa 1153
OY 1152 caatgttaacgataatcggttctacgcttccacagttcaacaatctgaactgtgt 1211
    |||||
DB 1154 caatgttaacgataatcggttctacgcttccacagttcaacaatctgaactgtgt 1213
OY 1212 tgccttcaactgtgtacaatcgtaacatcgtaacgcttcctcgcgactctggtgtgtctgt 1271
    |||||
DB 1214 tgccttcaactgtgtacaatcgtaacatcgtaacgcttcctcgcgactctggtgtgtctgt 1273
OY 1272 ggaatcatcccggtctgtgagcggtctgggggtgaacgttcgcgtgttaa 1317

```

```

DB 1274 ggaatcatcccggtctgtgacggttgggggtgaacgttcgcgtgttaa 1319
    |||||
RESULT 11
AAV30575
ID AAV30575 standard; DNA: 1546 BP.
XX
AC AAV30575;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C gene in phisBota.
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
XX botulism; ds.
XX
OS Clostridium botulinum serotype A.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT CDS 108..1496
FT /tag= a
XX
PN WO9808540-A1.
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US15394.
XX
PR 28-AUG-1996; 96US-0704159.
XX
PA (OPH1-) OPHIDIAN PHARM INC.
PI Thalleys BS, Williams JA.
XX
DR WPI: 1998-230234/20.
DR P-PSDB; AAM68390.
XX
PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
XX
PS Example 28; Page 277-278; 428pp; English.
XX
CC This is the DNA sequence of the Clostridium botulinum serotype A
CC toxin C fragment gene contained in plasmid phisBota. The encoded
CC toxin A polypeptide (see AAM68390) has a histidine-tagged
CC N-terminal extension. The vector was used to express native
CC (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC cells. The invention relates to recombinant proteins derived from
CC C. botulinum toxins. Methods are provided which allow for the
CC isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells.
CC The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
XX
SO Sequence 1546 BP; 629 A; 163 C; 256 G; 498 T; 0 other:

```

Query Match 55.0%; Score 727.6; DB 19; Length 1546;
 Best Local Similarity 72.2%; Pred. No. 1,2e-193;
 Matches 946; Conservative 0; Mismatches 364; Indels 0; Gaps 0;

```

OY 14 tgcctcactcatgaataatacaagaacatataactccatccctgaactcgcgt 73
    |||||
DB 193 tatctacattactgaataataataataataataataataataataataataataatgaat 252
OY 74 agcaatcaaacacccgatgacgtctgcgtacgcttccacaatcaaatcggttcca 133
    |||||
DB 253 atgaagaatcaatcaatgaatctatgattgattgattgattgattgattgattgattgatt 312

```


QY 134 aagttacatcgatccgagtcacaaagatccagatccagctgttcaatcctggaattcca 193
DB 313 aagtaaatcttgatccaatagataaaatccaatccaattatttatttagaagaatgta 372
QY 194 aaatcgaaagtatacctgaaagatgctacgtatatacaactctatgacgaaacttcca 253
DB 373 aaattgagtgtaatttaaaaaatgctatgtatataatagatgtaagtaaaatttagta 432
QY 254 cctcctcttgatccggtatcccgaaataactccaatccaatcctctcttgacaatgataca 313
DB 433 ctgactcttgatagaagaattccctaagtaatttaacagataagtaataatgataata 492
QY 314 ccaatcaactgacgtgaaacaaatctctgttggaagaatcctctgacacgaagtgtaa 373
DB 493 caataataatgtatgtaaaaaataatccaagatggaagatcaccttaattatggtgtaa 552
QY 374 tcaatcgtactctgagacacacacagaaatcaaacagcgtgtgtatccaatctctc 433
DB 553 taatcgtacttcaagatatactcaggaataaaacaaagtagtctttaataacagtc 612
QY 434 agatgatacaatcctctgactacatcaatcgctgctgttccgtaccacacaatc 493
DB 613 aaatgataatatacagatataataaacagatgattttgttaactatcactaataata 672
QY 494 gctgaataactccaacaaatctacacacggtcgtctgacgacgaagaaacagatctcca 553
DB 673 gattaaataacttaaatattatataaataagaaatgaagatgaatcaaaacaaatttcca 732
QY 554 atctggtgaacatccacgcttcaataacacatglttcaactgagcgtgttcgtgaca 613
DB 733 attaggaataatatactgtagtaataatgaatgtttaaatagatggtgttagagata 792
QY 614 ctacacgtacatctgtgatacaatcaatcgttgcgacaagaactgaaacgaagaag 673
DB 793 cactatagatatacttggaataaataatttaactcttcttgataaggaatlaaatagaagaag 852
QY 674 aaatcaaaagacctgacgaacacacgctccaatctctggtacccgtgaagactctgggtg 733
DB 853 aaatcaaaagattatataatgataatcaatcaatcaatcaggtatatttaaaagacttggggg 912
QY 734 actacgtcagtaacgacaacacgctacatgctgaactctgtacacgtcgaacaaatacag 793
DB 913 attattcaataatgataacacatcactatglttaacttattatgataatgaataatg 972
QY 794 ttgacgttaacaagttagtatacgcggttaccatgtaacgtccgcgtgttctg 853
DB 973 tgcagtataaataatgtaggtatagaggtatatagtatctttaaaggcctagaggtagcg 1032
QY 854 ttatgactacaacacatcactcgaactcttccctgtaccggtgtacacaaatccatcatca 913
DB 1033 taatgactacaacacatcattatattcaagtgttgaatgtaggggacaaatattatataa 1092
QY 914 agaaatacgcgtctgtaacaaagacataatcgttcgacaacatgactggtatataca 973
DB 1093 aaaaatacgtctcgtgaataaagaataatgttgaagaataatgatcgtgtatataata 1152
QY 974 atgttgtatgtaagaacaaagaataacgctcgtgcgtacacaaatgcttcaagcgtgttag 1033
DB 1153 atgttgtatgtaagaacaaagaataatgtaggtacgtacaaatgctacacagcagcgttag 1212
QY 1034 aaagaatctgtcgtcgtgaaatcccgagcgttgcgtgaactctgctcaggtaggtttaa 1093
DB 1213 aaaaaatcctaagtgcatgaataactcgtatgtaggaatcctaagtcgaagtgaagtaa 1272
QY 1094 tgaataccaagaacgacaggtgtacactaaacaaatgcaaaatgtaactgacgagcaaca 1153
DB 1273 tgaagtcaaaaaatgataagaataacaaataaataatgcaaaatgatttcaacgaagtaata 1332
QY 1154 atgtaacgatatcoggttctcgttccacacagttcaacaaatcgtcaacacgtgtg 1213
DB 1333 atgggaatgatataggtcttataagattcagttcaataataatagctaaacacagtag 1392

QY 1214 ctccaactggtacacatcgctacatcgacgtcttccctcgtcacctctggtgtctgtgg 1273
DB 1393 cagtaaatctggtatataatagacaataatgaagctcagtagagccttgggtctcaatgg 1452
QY 1274 agttaccccggtgtgacggtgtgggtgaacgtcgcgtctgaagaatc 1323
DB 1453 aatttattcctgtagatgtagatggtggggaagaagccactgaatataac 1502

RESULT 12

AAc64582
ID AAC64582 standard; cDNA; 4835 BP.
XX
AC AAC64582;
XX
DT 15-FEB-2001 (first entry)
XX
DE BONT/A neurotoxin prototoxin nucleotide sequence SEQ ID NO:10.
XX
KW Human; procholecystokinin; CCK A receptor; CCK B receptor;
KW pancreaticitis; antinflammatory; ss.
XX
OS Clostridium botulinum.
XX
PN WO200061192-A2.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09142.
XX
PR 08-APR-1999; 99US-0288326.
XX
PA (ALLIR) ALLERGAN SALES INC.
XX
PI Steward LE, Sachs G, Aoki KR.
XX
DR WPI; 2000-679416/66.
XX
PT New composition for treating acute pancreatitis, comprises a pancreatic cell surface marker binding element, a translocation element that transfers polypeptide across vesicular membrane, and a therapeutic element .
XX
PS
XX
Dislosure; Page 29-32; 50pp; English.
XX
CC The present invention describes a composition (I) for treating acute pancreatitis. (I) comprises a first element containing a binding element that binds to a pancreatic cell surface marker, a second element containing a translocation element that facilitates polypeptide transfer across the vesicular membrane, and a third element containing a therapeutic element that inhibits enzyme secretion in pancreatic cell cytoplasm. Also described is a method for making a therapeutic polypeptide having a binding element selective for cholecystokinin (CCK) receptor by expressing within a host cell a recombinant chimeric polypeptide comprising an extein containing a therapeutic element and a translocational element, and an intein located to the carboxy terminal of extein having a cysteine, serine or threonine at its amino terminus, and contacting the extein with a synthetic peptide comprising a CCK amino acid sequence containing an amidated phenylalanine at a natural C-terminus, and a cysteine, serine or threonine at its N-terminus, and a nucleophilic reagent able to cause cleavage of the intein to form a peptide bond between the extein C-terminus and synthetic peptide N-terminus through the formation of an activated ester or thio ester intermediate. (I) has antinflammatory activity and prevents accumulation of pancreatic digestive enzymes, and prevents exocytic fusion of vesicles containing secretory enzymes of pancreas. (I) is useful for treating acute pancreatitis. The present sequence encodes the Clostridium botulinum BONT/A neurotoxin prototoxin which is given in the exemplification of the present invention.
CC
XX
SQ Sequence 4835 BP; 1934 A; 517 C; 756 G; 1628 T; 0 other;

Query Match	55.0%;	Score 727.6;	DB 21;	Length 4835;
Best Local Similarity	72.2%;	Pred. No. 2e-193;		
Matches 946;	Conservative 0;	Mismatches 364;	Indels 0;	Gaps 0;

QY	14	tgcttacctcgaatgatatataagaacatcatcaatctccctccactcctgaaacctgacct	73
Db	2945	tattcaactcttaaccgaatatataataagaataattatttaattctctatctctatgtaattgaag	3004
QY	74	acgaatccaaatcacccfpatatgcacctctctcgcgtacagcttccaaatacaaatcgcgtctta	133
Db	3005	atgaaagttaacttaattaatgacttactctagtgatgcattcaaaaataaatatttggtagta	3064
QY	134	aagttaaacttcgatccgatctgcagacaagaatccagatccagctgtgtcaatctggaatctcca	193
Db	3065	aagtaaaattcttgatcccatgataaaaaatccaattccaattatttaatttgaagaagtta	3124
QY	194	aaatcgaagtatattccctgaagaatgctcatcgtatatacaactctatgtaagcaaaactctcca	253
Db	3125	aaattgaggccaattttaaaaaatgcatcgttatataatagtgtatgtaaaaaattttagta	3184
QY	254	ccctctcttgatccglatcccgaaatctcaactccatctctctctgacaatgataca	313
Db	3185	ctagcttcttgataagaatctccctaagtatattttaacagtaataagctcaataatgataata	3244
QY	314	ccatcatcaactgatatgaaaacaattcttggttgaaaatctctcttgaaactcgttgaa	373
Db	3245	caataaataaattgtaatgaaaataaattccggaggaagatcacacttaattatgtgtgaa	3304
QY	374	tcatctgagactctgcaggaacactcaggaatacacaagaagtttgtlatccaatctctc	433
Db	3305	taactcgtactttaaaggatcagatctcaggaataaacaagaagtagtttttaataatcagtc	3364
QY	434	agatgatccaacatctctcgtactatacaaalcgctcgtgactcttgatccatccaacaatc	493
Db	3365	aaatgattaaataatactcgaattataataacaagaatgttttcttaactcactaaataa	3424
QY	494	gtccgaataactccaanaatctatatacaagccgctctgtacaccagaagaacgagatctcca	553
Db	3425	gattaaataactctaaaattataataaayggaagattaatagatacaaaaaccaatttcaa	3484
QY	554	atctgggtaaacatccacgcttcttaataacaatcattgtccaactgtagcgtttgtctgta	613
Db	3485	attaggtaatatatcagctagtaataataatagttttaatttagatgtgtgtagagata	3544
QY	614	ctcgcgcctcatctgatatcaataatcttcaatctgttctgcacaagaacatcgaaacgaagaag	673
Db	3545	cacatgatactatcttgataaataatcttcaatcttcttcttgataagaattaaatgaaagaag	3604
QY	674	aaatcagaagacctgtagaacaaacgaatcccaattctgtatactctgaaagagatcttgggtg	733
Db	3605	aaatcagaagattatataatgataatcaatccaatccgaattcagtatatttaagaactttgggtg	3664
QY	734	actacctgcagtaagacaaacgctactacaatgctgaattctgtagatccgacaacaatacg	793
Db	3665	atttattacaatatgtaaaacatactataatgtttaaatttataatataatccaataataatg	3724
QY	794	ttgacgtccaacaatgttaggtgatccgcggtatgaactgaactcgtgaagaagtcgcggtgtctg	853
Db	3725	tcgtagtcaaatatgtctaggtatcattagaggttatatgtatattcctaagaagccttagaggtcag	3784
QY	854	ttatgactaccaaatctataccgtgaactctccctgtacggtgtgataccaatcatcatca	913
Db	3785	taatgactaccaaatcttatttatttaattccaagtgttatagggggacaattattattata	3844
QY	914	agaaatcgcgtctgtgtgaacaaggaacatatcgtctgcacaataatgatacgtgtatacatca	973
Db	3845	aaaaaattgctctcgtgnaataaagaataatattgttgagaataatgatacgtgtatacatca	3904
QY	974	atgtgtgagctttaaagaacaagaatctacgctcgtgcacaaatgctctccagcgtgtgtg	1033
Db	3905	atgtgtgagctttaaagaacaagaatataagtttagctctactaatgtcatccacagcgagcgtgag	3964
QY	1034	aaaaagatctgtcgtcctgynaatcccggaagctgtgttaatctgtctcaggtagttgttaa	1093

Dh	3965	aaaaatctactcaagtctgatttagaatacttcgtatgttagaattcaagtagtagta	4024
Qy	1094	tgaattccaagaacgacccaggtatcatcaacaatgcaaatgtatctcgagagcaaa	1153
Dh	4025	tgaagctaaaaaatgacaaaggaatacaataaagcaaaatgatcttacaagtaata	4084
Qy	1154	atgtaacgaatacgtgtttcatcgtgtttccaccagttcaacataatcgtaaacgtgtg	1213
Dh	4085	atggaaatgataagagcttatataggaattcatgaatttaataataaactaaacagtag	4144
Qy	1214	cttccaaactgtacaaatcgtcagatcgaacagttccctcgcacatctgtgtctcttg	1273
Dh	4145	caagtaattgtaataaagaacaaatagaagaactgtagagccttgggttcctcagtg	4204
Qy	1274	agttcatcccggtgttatgaaagctgtgggtgaaagctcgcgtctgtaagaatc	1323
Dh	4205	aatttattcctctgacagatgagatgggtgggaaagagccacttaactaaac	4254

	RESULT	13	
AAZ87220			
ID	AAZ87220	standard; DNA; 1317 BP.	
AC			
XX	AAZ87220;		
DT	08-MAY-2000	(first entry)	
XX			
DE	DNA encoding native BONT serotype A (BONTA) C-terminal fragment (Hc).		
XX			
KM	Botulinum neurotoxin; heavy chain; BoNT _H ; serotype A;		
KW	C-terminal fragment; Hc; Venezuelan equine encephalitis virus replicon;		
XX	VEE; botulism; vaccine; diagnosis; drug screening; ds.		
OS	Clostridium botulinum.		
FH	Key	Location/Qualifiers	
FT	CDS	1..1317	
FT		/tag= a	
FT		/product= "Native botulinum neurotoxin serotype A"	
FT		(BoNTA) heavy chain C-terminal fragment"	
FT		/transl_except= (pos:1294..1314, aa:cly)	
PN			
XX	WO200002524-A2.		
PD	20-JAN-2000.		
PF	09-JUL-1999;	99MO-US15570.	
XX	10-JUL-1998;	98US-0092416.	
PR	12-MAY-1999;	99US-0133870.	
XX	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.		
PA			
XN	Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;		
PJ	MPJ; 2000-160827/14.		
DR	P-PDB; AAY77142.		
PT	Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum toxin serotypes A-G, is used for inducing an immune response against botulinum -		
PS	Example 3; Page 52; 54pp; English.		
CC	The invention relates to novel vaccines that induce a protective immune response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F and G (BoNTA-BontG). The vaccine of the invention is novel recombinant DNA construct comprising a vector, and at least one nucleic acid fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT serotypes A-G. In preferred embodiments of the invention, the vector is a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this vector results in the production of large amounts of a protein encoded by		

a sequence cloned into the replicon. The constructs are used to produce CC vaccines against botulism. The proteins can also be used as diagnostic CC tools for the diagnosis of botulism. The transformed host cells can be CC used to analyse the effectiveness of drugs and agents which inhibit toxin CC effects. The vaccine currently used against botulism is dangerous CC and expensive to produce, and contains formalin, which is very painful CC for the recipient. Also, the vaccine is incomplete, in that only 5 of CC the 7 serotypes are represented in the formulation. The novel vaccine CC of overcomes these problems, as it is easily purified, and available in CC large quantities. It is also expressed in the lymph nodes for a better CC immune response. The present sequence represents DNA encoding native CC BONT/A heavy chain C-terminal fragment (Hc) used in an exemplification of CC the present invention.

Sequence 1317 BP; 560 A; 108 C; 211 G; 438 T; 0 other;

Query Match 54.9%; Score 726.4; DB 21; Length 1317;
Best Local Similarity 72.3%; Pred. No. 2,3e-193;
Matches 943; Conservative 0; Mismatches 361; Indels 0; Gaps 0;

QY 14 tcttacccttcgtatcatcaagaacatcatcaatccctccgaactcgtcgt 73
DB 14 tcttacccttcgtatcatcaagaacatcatcaatccctccgaactcgtcgt 73
QY 74 acgaatccaataccggtacgacctcgcgtacgctcccaaatccaatcgcgttca 133
DB 74 atgaagaatcatcaatcaatgaactctcagtgatgcacaaaataatcgtgagta 133
QY 134 aagtttaactcgtacgcgaacgaacagatcagatcgttcaactcgtgaacttcca 193
DB 134 aagtttaactcgtacgcgaacgaacagatcagatcgttcaactcgtgaacttcca 193
QY 194 aagtttaactcgtacgcgaacgaacagatcagatcgttcaactcgtgaacttcca 253
DB 194 aagtttaactcgtacgcgaacgaacagatcagatcgttcaactcgtgaacttcca 253
QY 254 cctccttcgtatcgcgtacgcgaacgaacatcctcctcgaacgaatgataca 313
DB 254 cctccttcgtatcgcgtacgcgaacgaacatcctcctcgaacgaatgataca 313
QY 314 ccatcaatcgtatcgtacgaacgaacatcgttgcgaacgaatcgttgcgaac 373
DB 314 ccatcaatcgtatcgtacgaacgaacatcgttgcgaacgaatcgttgcgaac 373
QY 374 tcatcgtacgtcgtcgaacgaacgaacgaacgaacgaacgaacgaacgaac 433
DB 374 tcatcgtacgtcgtcgaacgaacgaacgaacgaacgaacgaacgaacgaac 433
QY 434 agatgatacaatcctcgtacgaacgaacgaacgaacgaacgaacgaacgaac 493
DB 434 agatgatacaatcctcgtacgaacgaacgaacgaacgaacgaacgaacgaac 493
QY 494 gttcgtgaataatcctcgtacgaacgaacgaacgaacgaacgaacgaacgaac 553
DB 494 gttcgtgaataatcctcgtacgaacgaacgaacgaacgaacgaacgaacgaac 553
QY 554 atctggtgaataatcctcgtacgaacgaacgaacgaacgaacgaacgaacgaac 613
DB 554 atctggtgaataatcctcgtacgaacgaacgaacgaacgaacgaacgaacgaac 613
QY 614 ctcacgcgtacatcgtacgaacgaacgaacgaacgaacgaacgaacgaacgaac 673
DB 614 ctcacgcgtacatcgtacgaacgaacgaacgaacgaacgaacgaacgaacgaac 673
QY 674 aaatcaagaacgtgatacgaacgaacgaacgaacgaacgaacgaacgaacgaac 733
DB 674 aaatcaagaacgtgatacgaacgaacgaacgaacgaacgaacgaacgaacgaac 733
QY 734 actaactgagtaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 793
DB 734 actaactgagtaacgaacgaacgaacgaacgaacgaacgaacgaacgaacgaac 793

QY 794 ttgacgtcaacaatgtagatccgcggttaccatgtagtaccgaagtcgcggtgttcg 853
DB 794 ttgacgtcaacaatgtagatccgcggttaccatgtagtaccgaagtcgcggtgttcg 853
QY 854 ttatgataccaatcctcgtacgaacgaacgaacgaacgaacgaacgaacgaac 913
DB 854 ttatgataccaatcctcgtacgaacgaacgaacgaacgaacgaacgaacgaac 913
QY 914 agaaatgagtgtagtgaagaagaagaagaagaagaagaagaagaagaagaaga 973
DB 914 agaaatgagtgtagtgaagaagaagaagaagaagaagaagaagaagaagaaga 973
QY 974 atgtgtagtgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1033
DB 974 atgtgtagtgaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1033
QY 1034 aaaagatctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1093
DB 1034 aaaagatctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1093
QY 1094 tgaatccaagaacgagcagcagcagcagcagcagcagcagcagcagcagc 1153
DB 1094 tgaatccaagaacgagcagcagcagcagcagcagcagcagcagcagcagc 1153
QY 1154 atgtgaacgaatcgttccatcgttccatcgttccatcgttccatcgttccatc 1213
DB 1154 atgtgaacgaatcgttccatcgttccatcgttccatcgttccatcgttccatc 1213
QY 1214 ctcccaactgtagaactcgtcgaacgaacgaacgaacgaacgaacgaacgaac 1273
DB 1214 ctcccaactgtagaactcgtcgaacgaacgaacgaacgaacgaacgaacgaac 1273
QY 1274 agttccctcgtgtagtgaagcgttgcggtgtagtgaagcgttgcggtgtag 1317
DB 1274 agttccctcgtgtagtgaagcgttgcggtgtagtgaagcgttgcggtgtag 1317

RESULT 14

AA87218
ID AA87218 standard; DNA; 2532 BP.

AC AA87218;
XX

DT 08-MAY-2000 (First entry)
XX

DE DNA encoding native botulinum neurotoxin serotype A (BONT/A).
XX

KW Botulinum neurotoxin; heavy chain; BONT; serotype A;
KW Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening; ds.
XX

OS Clostridium botulinum.
XX

FH Key Location/Qualifiers
FT 1..2532
FT /tag= a
FT /product= "Native botulinum neurotoxin serotype A
FT /transl_except= (pos:2509..2529, aa:GLY)
FT

PN MO200002524-A2.
XX

PD 20-JAN-2000.
XX

XX 09-JUL-1999; 99MO-US15570.
XX

XX 10-JUL-1998; 98US-0092416.
XX

PR 12-MAY-1999; 99US-0133870.
XX

PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX

XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L,
PI

DR MPI: 2000-160827/14.
 DR P-PSDB: AAT77140.
 XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
 PT toxin serotypes A-G, is used for inducing an immune response against
 PT botulinum -
 PS Example 3: Page 49-50; 54pp: English.

XX The invention relates to novel vaccines that induce a protective immune
 CC response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
 CC and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
 CC DNA construct comprising a vector, and at least one nucleic acid
 CC fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
 CC serotypes A-G. In preferred embodiments of the invention, the vector is a
 CC Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this
 CC vector results in the production of large amounts of a protein encoded by
 CC a sequence cloned into the replicon. The constructs are used to produce
 CC vaccines against botulinism. The proteins can also be used as diagnostic
 CC tools for the diagnosis of botulinism. The transformed host cells can be
 CC used to analyse the effectiveness of drugs and agents which inhibit toxin
 CC effects. The vaccine currently used against botulinism is dangerous
 CC and expensive to produce, and contains formalin, which is very painful
 CC for the recipient. Also, the vaccine is incomplete, in that only 5 of
 CC the 7 serotypes are represented in the formulation. The novel vaccine
 CC of overcomes these problems, as it is easily purified, and available in
 CC large quantities. It is also expressed in the lymph nodes for a better
 CC immune response. The present sequence represents DNA encoding native
 CC BoNTA heavy chain used in an exemplification of the present invention.
 XX
 XX Sequence 2532 BP; 1038 A; 241 C; 404 G; 849 T; 0 other;

Query Match 54.9%; Score 726.4; DB 21; Length 2532;
 Best Local Similarity 72.3%; Pred. No. 3.2e-193;
 Matches 943; Conservative 0; Mismatches 361; Indels 0; Gaps 0;

QY 14 tgcctactcactgaatcatcatcaagaacatcatcaatcctcctgaacgtgcgt 73
 DB 1229 tactcatcattactgataatataatgaatatttaatacttcataatgattgaat 1288
 QY 74 agcaatccaatcccgatgcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 133
 DB 1289 atgaaagtaattcatttaatactgactcgtatgcatgcatcaaaaataaatctgtgtgta 1348
 QY 134 aagtlactcgtatccgatcgacagaatcagaatccagctgttcactcgtgaactctcca 193
 DB 1349 aagtaattttgtatccaatatgataaaaatacaattcaattatatttaagaagtgtga 1408
 QY 194 aatcgaagtatcccgatcgacagaatcagaatccagctgttcactcgtgaactctcca 253
 DB 1409 aattgaggttaatttaaaaaatgctatgtatataatagtagtatagtagaatatttagta 1468
 QY 254 cctccttcgtgacgtcgtatccgaatcctgaactcctcctcctcgtgaacatgaatata 313
 DB 1469 ctgacttttggataaagaatcccaagtattttaaagtaagaatccttaataaagaatata 1528
 QY 314 ccatcatcaactgcatggaanaacatctggttggaaagtatcctgaactcgtgaatga 373
 DB 1529 caataataatgtatggaanaaataatcagatggaagtaactcaacttaattatgtgaa 1588
 QY 374 tcatctggactcgcaggaactcaggaataacacagcgtgtgtgtatccaactctc 433
 DB 1589 taactcggactctacaggaactcaggaataaacaagaagtagtattttaataacagtc 1648
 QY 434 agatgcaacatctcgtgatacatcaatcgtcgtggtatcgttaccatcaacaactc 493
 DB 1649 aatatgattatatacatgataataaacaagatgattttgtaaccataactaataata 1708
 QY 494 gtcgtgataactccaataatcatcaacgacgctcgtgacgcgaacacacgatactcca 553
 DB 1709 gatataaactctataaattatataataatggaagattatagatcaaaaacaaatttcaa 1768

QY 554 atctgggtaaacatccacgcttctaataacatcatgttccaacgtgactgtgtaca 613
 DB 1769 attaggtaaatcatcagctcagtagtaataataatgatttaattagatgtgtgagata 1828
 QY 614 ctacccgtcatctcgtgataaataacttaactcgttcgcgaagaactcgaagaag 673
 DB 1829 caatagatatacttgatataaataatatttaactcttttgaataagaactaagaagaag 1888
 QY 674 aaatcaagaacgtgtaacgaacacagctcccaattcgtgatactcgtgaagaactcgtggt 733
 DB 1889 aaatcaagaactttatagataatcaatcaatcaatcaggtatttaagaagactttgggggt 1948
 QY 734 actacgtcgtacgaacaaacggtacatcagatgctgaactcgtgacgtcgaacaaatagc 793
 DB 1949 attattaccatatagtataaaccataactatgtttaattatatagtaccacaataaatg 2008
 QY 794 ttgacgtcaacaatgtatggtatccggtgtacatgtgtcccgaaaggtccggtgtgtcgt 853
 DB 2009 tcatgtataaataatgtagtattagaggttatatagcttaaaaggccatagaggtagcg 2068
 QY 854 ttatgactacaacatctcactcgtgaactctcccgtaacgtgtgataccaatcatca 913
 DB 2069 taatgactacaacatcttatttaataatcagaatgtgtataggggacaaatattatata 2128
 QY 914 agaaatacgcgtctgtgtacaagaacataatcgttcgcaacaatgatcgtgtataatca 973
 DB 2129 aaaaatacgtctcgtgaataaagaatataatgtttagaataatgaltcgtgtatata 2188
 QY 974 atgttgaagttaagaacaaagaataccgtctgtgtacaaatgcttcgaagcgtgtgtag 1033
 DB 2189 atgtagtagtataaataaagaataataggttagtactaaatgctccaggtcaggtcag 2248
 QY 1034 aaaagatctgtcgtcgtcgtgaaatcccgagcgtgtgttaactgtctcgaagtagtgtaa 1093
 DB 2249 aaaaatacctaacgagcttagaataatcactgattgtaggaatctcaagtagtagtaaa 2308
 QY 1094 tgaatccaaagaacgacaggtgtatcactaaacaaatgcaaatgaaatcgtcaggaacaa 1153
 DB 2309 tgaagtcataaataatgcaagaataaataaataatgcaaatgaaatgaaatgaaatgaa 2368
 QY 1154 atgttaacgatatcgtgttcacgtgttcacaggttcacaaatcagtaactcgtgtg 1213
 DB 2369 atggaatgatataggtcttataatgattcactcaggttttaataatataatgcaaaactag 2428
 QY 1214 ctcccaactgtgtacatcgtcagatcgaacgttccctcgtcactcgtgtgtcgtctgg 1273
 DB 2429 caagtaattgtatataatagcaaatagaaagatcctagtaggacttgggtgtcgtcatgg 2488
 QY 1274 agttcatcccggttgatgacggttgggtgtgaaggtccgctgttaa 1317
 DB 2489 aatttattcctgtgatatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 2532

RESULT 15
 AAT29244
 ID AAT29244 standard; DNA; 3891 BP.
 XX
 AC AAT29244;
 XX
 DT 09-JUL-1996 (first entry)
 XX
 XX C. botulinum type A neurotoxin gene.
 DE
 DE Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen; ds.
 KW
 OS Clostridium botulinum.
 XX
 PN WO9612802-A1.
 XX
 PD 02-MAY-1996.
 XX
 PF 23-OCT-1995; 95MO-US13737.
 XX

```

PR 07-JUN-1995; 95US-0480604.
PR 24-OCT-1994; 94US-0329154.
PR 16-MAR-1995; 95US-0405496.
PR 14-APR-1995; 95US-0422711.
XX
XX (OPHI-) OPHIDIAN PHARM INC.
PA
XX Firca JR, Klink JA, Padhye NV, Stafford DC, Thalley BS;
PI Williams JA;
XX WPI: 1996-230603/23.
DR P-PSDB: AAR95010.
XX
XX Fusion proteins comprising non-toxin protein and part of toxin
XX useful to form anti-toxins against Clostridium botulinum type A, and
XX C. difficile type toxins, and to treat C. difficile intoxication,
XX partic. diarrhoea
XX
XX Example 22; Page 344-350; 434pp; English.
XX
XX The Clostridium botulinum type A neurotoxin gene (AAR29244) has been
XX cloned and sequenced. It encodes a protein (AAR95010) that is
XX processed to form a dimer composed of a light and a heavy chain.
XX The 50 kDa C-terminal portion of the heavy chain, or C fragment
XX (AAR95008), was produced using a synthetic gene (AAR29245) having codon
XX usage altered to improve expression in Escherichia coli. Fusion
XX proteins of the type A toxin or C fragment, with e.g. maltose
XX binding protein or polyhistidine affinity tag, are used to generate
XX neutralising antitoxins and in vaccine compsns.
XX
XX Sequence 3891 BP; 1580 A; 392 C; 629 G; 1290 T; 0 other;
SQ

```

Query Match 54.9%; Score 726.4; DB 17; Length 3891;
 Best Local Similarity 72.3%; Pred. No. 3.9e-193;
 Matches 943; Conservative 0; Mismatches 361; Indels 0; Gaps 0;

```

QY 14 tgcctaccttcgaatcacatcaagaacatcatcaatccctccgaactgcgt 73
DB 2588 tatctacatttactgaatatataatgaatatattactctctatattgaattgaat 2647
QY 74 acgaatcaatcacctgatcgactgcgtcgaactcccaaatcaatcagttctta 133
DB 2648 atgaagaatattatattatagacttactaggtatgcacaaataaataatgtagta 2707
QY 134 aagtaactgcgatcgatcgacgaagaatcagatccagctgttcaatctggaatcttca 193
DB 2708 aagtaaatcttgcataccaatagataaaatcaatccaattcaatttaattagaagaatgta 2767
QY 194 aatcgaaagttaacctgaagaatgctatcgtatcacactctatgtacgaataactctca 253
DB 2768 aatcgaggttaattttaaataatgctatgtatataatgtatgtatgtatgtatgtatgtat 2827
QY 254 cctcctctgtgactcgtatcccgaaatattcctaactcactctctctgaaatgtaataca 313
DB 2828 ctgagcttctgtgataagaatcctcctaagatttcaacagtaatacgttaaatatgaata 2887
QY 314 ccatatcaactgcgatgtaaaacaaattctgtgtgaaagtatctctgaactcagtgtaaa 373
DB 2888 caataataatgtatgtaaaataatcagagatgtaaaatcactcaacttaattgtgtgaaa 2947
QY 374 tcaatgtgactgcgaagacatcaggaataatcaaacagcgtgtgtatccaataactctc 433
DB 2948 taatctggaacttacaagatctcaggaataaatacaagaatgagtttttaataacagtc 3007.
QY 434 agatgatcaaacatctctgtacatcaatcagtcgtgtatctgtttacatccacaacatc 493
DB 3008 aatgatataataatcagattatatacaacagatgagttttgttaactacactaataca 3067
QY 494 gctgaataataactccaaataactcaatcaacgagcgtgtgtatccagaacacgatactca 553
DB 3068 gattaataactcctaataattatataaataatgtaagaatlaataagatacaaaaacaaatttcaa 3127

```

```

QY 554 atctggtaacatccacgcttctcaataacatcatgttcaaacatgtagcggtgtgtcgtgaca 613
DB 3128 atttagtaataatctatcgtctagtaataataatgtaataatgtaattagttgtgtgagata 3187
QY 614 ctacccgtacatctgtgataccaatattcctaactgtgttcgacaagaatgtagcaaaaag 673
DB 3188 cacatagataatattgtataaaatatttcaatctcttcttgataaggaatttaaatgtaaaaag 3247
QY 674 aatacaagaacctgtgacgaacacgaatcccaattctgtgtatccgaaagcttctggggtg 733
DB 3248 aatacaagaattatataatgaataatcaatcaatcaggtatctttaaagaacttttgagggtg 3307
QY 734 actacctgagtaagcaacaacgtaatacgtatcgtatcgtatcgtatccgaacaataacg 793
DB 3308 attattacaatctgtataaacacatactatagttaattatattatgtatccaaataaataag 3367
QY 794 tgaagtcacaacatgtagtatccggttaccgtgtacatcgtgaagaagtcgcgtgttctg 833
DB 3368 tcgattgaataatgtagtatgaggtatagatatatctcttaaaaggcctagaggtagcg 3427
QY 854 ttatgactcaacacatctacactcctgaactctccctgtgacggtgtgtacaaatcatcatca 913
DB 3428 taatgactcaacaacatttatttaattcaagttgtgtataggggacaacaattatataaa 3487
QY 914 agaaatcagcgtctgtgtaacaagaagaataatcgttgcgaacaatgtatcgtgtatataca 973
DB 3488 aaaaatattctcttggaataaagaataatattgttggaataatgtagtgcgtatataat 3547
QY 974 atgtgtagttagaagaacaagaatccggtctgctcacaatgtcttccaggtgtgtgtag 1033
DB 3548 atgtagtagttaaaaaataagaatatagtttagtactactaatgcatcacagcgagcgtag 3607
QY 1034 aaaaagcttctgtcgtctggaataatcccgagcgtgtgtatctgtctcaggtatgtgttaa 1093
DB 3608 aaaaatactaagtgtcattagaataatcccgatgtgaggaatctcaagttcaagttgtagtaa 3667
QY 1094 tgaatccaaagaacgacaggtgtatcactaacaatgtcaaaatgtatctgtcgaagacaaca 1153
DB 3668 tgaagtcaaaaatgtatcaaggaataacaataatgtcaaaaatgtatcttcaagaataata 3727
QY 1154 atgtgaagatatacgtttcatctcgtgttccacaggttcaacaatatacgtctaaactgtgtg 1213
DB 3728 atggaatgatataggtcttataagatttcatcagtttaataatagctaaactagtag 3787
QY 1214 ctccacaggtgtacaaatcgtcagatcgaacgttccctcctgcagctcgtggtgtctgtgag 1273
DB 3788 caagtaattcgtataatagacaataatgaaagatcctagtagacttcggtgtctcatgag 3847
QY 1274 agttacatcccggtgtgtatgacggtgtggtgtgaaacgttcggtgttaa 1317
DB 3848 aatttattcctgtagatgtagtggatggggaagaagccacatgttaa 3891

```

Search completed: September 16, 2002, 21:43:19
 Job time: 12921 sec

• • • • •

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 21:32:06 ; Search time 139.08 Seconds
(without alignments)
2336.591 Million cell updates/sec

Title: US-09-611-419A-3

Perfect score: 1323

Sequence: 1 gaatcgaaacgagtcctac.....aacgtccgtgtaagaattc 1323

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/PTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/Dackfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1304.4	98.6	1330	1	US-08-480-604A-22
2	1304.4	98.6	1330	2	US-08-405-496A-22
3	1304.4	98.6	1330	4	US-08-915-136-22
4	1304.4	98.6	1402	1	US-08-480-604A-25
5	1304.4	98.6	1402	2	US-08-405-496A-25
6	1304.4	98.6	1402	4	US-08-915-136-25
7	726.4	54.9	3891	1	US-08-480-604A-27
8	726.4	54.9	3891	2	US-08-405-496A-27
9	726.4	54.9	3891	4	US-08-915-136-27
10	166.4	12.6	1359	1	US-07-618-312A-3
11	166.4	12.6	1359	1	US-08-280-228-3
12	74.2	5.6	1359	1	US-07-618-312A-1
13	74.2	5.6	1359	1	US-08-110-786A-7
14	74.2	5.6	1359	1	US-08-280-228-1
15	74.2	5.6	1359	1	US-08-668-381A-6
16	60	4.5	5511	3	US-08-928-361B-2
17	60	4.5	7334	3	US-08-928-361B-1
18	57	4.3	5163	3	US-08-700-651-1
19	57	4.3	5163	3	US-08-928-361B-4
20	57	4.3	518	3	US-08-700-651-2
21	57	4.3	5318	3	US-08-928-361B-3
22	40	3.0	1690	1	US-08-276-452A-24
23	40	3.0	1690	2	US-08-798-744-24
24	37.8	2.9	2277	1	US-08-676-967-2
25	37.8	2.9	2277	1	US-08-676-974-2
26	37.8	2.9	2277	2	US-09-098-487-2
27	37.2	2.8	18596	4	US-09-318-448-11

28	36.8	2.8	1430	1	US-08-276-452A-25	Sequence 25, Appl
29	36.8	2.8	1430	2	US-08-798-744-25	Sequence 25, Appl
30	34.6	2.6	2094	4	US-08-714-918-87	Sequence 87, Appl
31	34.6	2.6	2094	4	US-09-265-315-87	Sequence 87, Appl
32	34.6	2.6	2094	4	US-09-265-315-87	Sequence 87, Appl
33	34.6	2.6	2094	4	US-09-266-417-87	Sequence 87, Appl
34	34.4	2.6	951	1	US-07-855-412B-2	Sequence 2, Appl1
35	34.4	2.6	951	2	US-08-308-887A-2	Sequence 2, Appl1
36	34.4	2.6	951	3	US-08-881-094-2	Sequence 2, Appl1
37	34.2	2.6	198	5	PCT-US95-10668-1	Sequence 1, Appl1
38	34.2	2.6	198	5	PCT-US95-10668-2	Sequence 2, Appl1
39	34.2	2.6	198	5	PCT-US95-10668-3	Sequence 3, Appl1
40	34.2	2.6	198	5	PCT-US95-10668-4	Sequence 4, Appl1
41	34.2	2.6	3292	1	US-07-814-964-12	Sequence 12, Appl
42	34.2	2.6	3292	1	US-08-258-442-12	Sequence 12, Appl
43	34.2	2.6	3292	1	US-08-328-809-7	Sequence 7, Appl1
44	34.2	2.6	3292	5	PCT-US92-11107-12	Sequence 12, Appl1
45	34	2.6	1666	4	US-09-134-607A-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-480-604A-22
; Sequence 22, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STEFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLTA, DIANE E.
; REGISTRATION NUMBER: 40,027

```

REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1314
US-08-480-604A-22

```

```

Query Match      98.6%; Score 1304.4; DB 1; Length 1330;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 12 gatgtacacctcaatgaatatacaagaacatcatcaactccatccatgaacctg 71
DB 12 gctgtaccttcactgaatatacaagaacatcatcaactccatccatgaacctg 71
QY 72 ctgaagaatccaatcaactgaatgaatgaatgaatgaatgaatgaatgaatga 131
DB 72 ctgaagaatccaatcaactgaatgaatgaatgaatgaatgaatgaatgaatga 131
QY 132 taagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 191
DB 132 taagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 191
QY 192 caaatacgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 251
DB 192 caaatacgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 251
QY 252 caactcccttctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 311
DB 252 caactcccttctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 311
QY 312 caactcccttctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 371
DB 312 caactcccttctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 371
QY 372 aatcatctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 431
DB 372 aatcatctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 431
QY 432 tcaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 491
DB 432 tcaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 491
QY 492 tcaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 551
DB 492 tcaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 551
QY 552 caatctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 611
DB 552 caatctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 611
QY 612 caatctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 671
DB 612 caatctggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 671
QY 672 agaaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 731
DB 672 agaaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 731
QY 732 tgactacatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 791
DB 732 tgactacatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 791

```

```

QY 792 cgttgacgtcaacaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 851
DB 792 cgttgacgtcaacaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 851
QY 852 tcttgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 911
DB 852 tcttgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 911
QY 912 caagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 971
DB 912 caagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 971
QY 972 caatgttgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1031
DB 972 caatgttgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1031
QY 1032 agaaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1091
DB 1032 agaaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1091
QY 1092 aatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1151
DB 1092 aatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1151
QY 1152 caatgttgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1211
DB 1152 caatgttgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1211
QY 1212 tcttgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1271
DB 1212 tcttgatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1271
QY 1272 ggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1317
DB 1272 ggaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaat 1317

RESULT 2
US-08-405-496A-22
; Sequence 22, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; NUMBER OF INVENTIONS: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405,496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791

```


PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPMD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1314
US-08-915-136-22

Query Match 98.6%; Score 1304.4; DB 4; Length 1330;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 12 gctgtaccttcaatgaatatacaagaacatcatcaatcaatcctgaactg 71
DB 12 gctgtaccttcaatgaatatacaagaacatcatcaatcaatcctgaactg 71
QY 72 ctgaagaatcaatcactgactgactgactgactgactgactgactgactg 131
DB 72 ctgaagaatcaatcactgactgactgactgactgactgactgactgactg 131
QY 132 taagaatcaatcactgactgactgactgactgactgactgactgactg 191
DB 132 taagaatcaatcactgactgactgactgactgactgactgactgactg 191
QY 192 caaatacgaagtatacctgaagaatgctatcgtatatacactcatgta 251
DB 192 caaatacgaagtatacctgaagaatgctatcgtatatacactcatgta 251
QY 252 caactcctcctggaatccgaataactcaactcactcctcctgaaga 311
DB 252 caactcctcctggaatccgaataactcaactcactcctcctgaaga 311
QY 312 caactcctcctggaatccgaataactcctggttggaagtatctctga 371
DB 312 caactcctcctggaatccgaataactcctggttggaagtatctctga 371
QY 372 aatcactggaactcgaagacactcaggaatacaacagcggtgtatca 431
DB 372 aatcactggaactcgaagacactcaggaatacaacagcggtgtatca 431
QY 432 tcagatgatacaatcctgactcaatcaatcactgactgactgactgact 491
DB 432 tcagatgatacaatcctgactcaatcaatcactgactgactgactgact 491
QY 492 tcgtcgaataactcgaataactcaatcaatcactgactgactgactgact 551
DB 492 tcgtcgaataactcgaataactcaatcaatcactgactgactgactgact 551

DB 492 tcgtcgaataactcgaataactcaatcaatcactgactgactgactgact 551
QY 552 caactcctcctggaatccgaataactcaactcactcctcctgaaga 611
DB 552 caactcctcctggaatccgaataactcaactcactcctcctgaaga 611
QY 612 caactcctcctggaatccgaataactcaactcactcctcctgaaga 671
DB 612 caactcctcctggaatccgaataactcaactcactcctcctgaaga 671
QY 672 agaatcgaagaactcctgactgaagaacaggtcccaatcctgactc 731
DB 672 agaatcgaagaactcctgactgaagaacaggtcccaatcctgactc 731
QY 732 tgactcctcctggaatccgaataactcaactcactcctcctgaaga 791
DB 732 tgactcctcctggaatccgaataactcaactcactcctcctgaaga 791
QY 792 cgttgaactcgaagaatcctgactgaagaacaggtcccaatcctgactc 851
DB 792 cgttgaactcgaagaatcctgactgaagaacaggtcccaatcctgactc 851
QY 852 tgatgaactcgaagaatcctgactgaagaacaggtcccaatcctgactc 911
DB 852 tgatgaactcgaagaatcctgactgaagaacaggtcccaatcctgactc 911
QY 912 caagaatacagcgtcgtgactgaagaacaggtcccaatcctgactc 971
DB 912 caagaatacagcgtcgtgactgaagaacaggtcccaatcctgactc 971
QY 972 caatggtgaatcgaagaatcctgactgaagaacaggtcccaatcctgactc 1031
DB 972 caatggtgaatcgaagaatcctgactgaagaacaggtcccaatcctgactc 1031
QY 1032 agaatcgaagaactcctgactgaagaacaggtcccaatcctgactc 1091
DB 1032 agaatcgaagaactcctgactgaagaacaggtcccaatcctgactc 1091
QY 1092 aatgaatacgaagaactcctgactgaagaacaggtcccaatcctgactc 1151
DB 1092 aatgaatacgaagaactcctgactgaagaacaggtcccaatcctgactc 1151
QY 1152 caatggtgaatcgaagaatcctgactgaagaacaggtcccaatcctgactc 1211
DB 1152 caatggtgaatcgaagaatcctgactgaagaacaggtcccaatcctgactc 1211
QY 1212 tgcttccaaactgactgaagaatcctgactgaagaacaggtcccaatcctgactc 1271
DB 1212 tgcttccaaactgactgaagaatcctgactgaagaacaggtcccaatcctgactc 1271
QY 1272 ggaatcgaagaactcctgactgaagaacaggtcccaatcctgactc 1317
DB 1272 ggaatcgaagaactcctgactgaagaacaggtcccaatcctgactc 1317

RESULT 4
US-08-480-604A-25
Sequence 25, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA

```

COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1386
US-08-480-604A-25

```

```

Query Match          98.6%; Score 1304.4; DB 1; Length 1402;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 312 caacatcatcaactgcatgaaacaattcgtgtgaaagtatctgaactgagtg 371
DB 384 CACCATCATCAACTGCGATGGAAAAACAAATTCGTGGTGGAAAGATCTCTCACTACGGTGA 443
QY 372 aatcatctggaactctgcaagacatcagaanaatcaaacagcgtgtgtatcaaatctc 431
DB 444 AATCATCTGGACTCTGCAGACACTCAGGAATCAAAACAGCGTGTATTCAAATATCTC 503
QY 432 taagatgatacaatctctgactataatcgtcgtgagatcttcgtttaccataccaacaa 491
DB 504 TCAGATGATCAACATCTCTGACTATCAATCAATGCTGGATCTTCTGTTACATATCAACAA 563
QY 492 tcgtctgaaatactccaatatcatcaacagcgctcgtgacgcagaacaacggatctc 551
DB 564 TCGTCTGAATTAATCTCAAAATCTACTACTCAACGGCGCTGTGATCGACCAAAACGATCTC 623
QY 552 caatctgggtaacatccacgcttctaataacatcatttcaaacctgagcgtgtgtcgtga 611
DB 624 CAATCTGGGTAAACATCCACGCTTCTAATTAACATCATGTTCAAACTGGACGGTGTCTGTA 683
QY 612 caatcagcgtacatctggtatcaaatctcaatctcgttcggaacaagaactgaagaaaa 671
DB 684 CACTCACCGCTACATCTGATGATCAAAATCTTCAATCTGTTGCAAAAGAACTGAACGAAAA 743
QY 672 agaaatcaaaagcctgtacgacaaccagtcacaattctgtatctcctgaagactctg99g 731
DB 744 AGAAATCAAAAGCCTGTACGACACACCAACGTCATTTGTGTATCTCAAAAGACTCTCGGG 803
QY 732 tgactacctcgaagtlacgacaaacgtaactacatgctgaatctgtacgacatcgaaacata 791
DB 804 TCACCTACCTGCGAGTACGACAAACCGTACTACATGCTGATCTGTACGATCCGAACAAATA 863
QY 792 cgttgaagcgaacatgtagtatacgcggttataatgataatccctggaagcgcggtgttc 851
DB 864 CGTTGACGTCAACATGTAGTATCCGCGGTATCATGTACTCTGAAGGTCGCGGTGTGTT 923
QY 852 tgtatgactacaacatcatctgaactcttccctgtaacgtgtgacccaatitacat 911
DB 924 TGTATGACTACCAACATCTACTACTGTAACCTTCCCTGTACCGTGTATCCAAATTCATCAT 983
QY 912 caaagaatacgcgctcgtgttaacaagaacatatcgttcgcaacaatgatacgtgtatcat 971
DB 984 CAAGAAATACGCGCTCTGTGTACAGGACAAATATCGTTCACAAATGATCGTATATACAT 1043
QY 972 caatgttgaattaaagaacaagaataacgcttgcgtacaaatgcttccaaagcgtgtgt 1031
DB 1044 CAATGTGTAGTTAAGAACAAAGAAATACCGCTCTGGCTACCAATGCTTCTAGGCTGTGT 1103
QY 1032 agaaaagatctgtctcgtctgtgaaatcccggaacgttgttaatctgtctcaggtagtgt 1091
DB 1104 AGAAAGATCTGTGTCTGTCTGTGAATCCCGGACGTTGGTAATCTGTCTCAGGTACTTGT 1163
QY 1092 aatgaataccaaagaacgaccaggtgtatcaactaaatgcaaaatgaatctgcaggaa 1151
DB 1164 AATGAATATCAAGAAAGACAGGATGATCACTAAATCAATCAAAATGAATCTGAGACAA 1223
QY 1152 caatgttaacgatatcgtgttcaatcggttcacacaggttcaaaatatcgtctaaactgt 1211
DB 1224 CAATGTATACGATATGCGTTTCAATCGGTTTCCACCGTTCACAAATATGCTTAATCGGT 1283
QY 1212 tgcctcacaactgtatacatcgttcagaatcgaacgttccctcgtcactctggtgtgtctgt 1271
DB 1284 TGGTTTCAACCTGTATCAATTCGATCAATGATCAAGCTTCTCTCGGACCTCTGGGTTGCTCTTG 1343
QY 1272 ggaatcaccacggtgtgatacgaaggttggtggtggaagtcgcgtgttaa 1317
DB 1344 GGAGTTCATCCGCGTGTGATGACGCTTGGGGTGAAGTCCGCTGTAA 1389

```

```

RESULT 5
US-08-405-496A-25
; Sequence 25, Application US/08405496A
; Patent No. 5919665

```



```

US-08-915-136-25
: Sequence 25, Application US/08915136
: Patent No. 6290960
: GENERAL INFORMATION:
: APPLICANT: KINK, JOHN A.
: APPLICANT: THALLEY, BRUCE S.
: APPLICANT: PADHIE, NISHA V.
: APPLICANT: FIRCH, JOSEPH R.
: APPLICANT: STAEFORD, DOUGLAS C.
: TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL, LLP
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/915,136
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/480,604
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/405,496
: FILING DATE: 16-MAR-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/329,154
: FILING DATE: 25-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/161,907
: FILING DATE: 02-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/985,321
: FILING DATE: 04-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/429,791
: FILING DATE: 31-OCT-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: INCOLIA, DIANE E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: OPHD-01763
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 25:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1402 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1386
: US-08-915-136-25

```

D	b	84	GCCTCTACTCTTCACGTGAATACATACATGAAAGAAATCAATCAATTAACCTTCATCTCTGAACCTGCG	143
Q	y	72	ctacgaatccaatcaacctgtaacgcactgtctcgtactacgcttccaaatacaatcagtttc	131
D	b	144	CTAGAGATCCAAATCACCTGTACGACCTGTCTGCGTACGCTTCGTAATAATCAACATGCGTTC	203
Q	y	132	taaaagtaacttgatccggtctgcaagaatcaagatccagcgtgttccaatcttgaaatttc	191
D	b	204	TAAAGTTAACTTCGATCCGATCCGATCAAGAATCAAGATCCAGCTGTTCAATCTGGAAATTC	263
Q	y	192	caaaatcgaagttacccctgaagaatgcatacgtatacaactatgtacgaanaatttc	251
D	b	264	CAAAATGGAAGTATCTCTGAAAGATGCTATCGTATCAACTCTATGTACGAAAACTTCTC	323
Q	y	252	caccctctcttgatccgtatcccgaaatacttcaactccatctctcttgaaatgaata	311
D	b	324	CACCTCTCTTGATCCGTCGTCGCCGAAATTACTCAACTCACTCTCTGAACAATGAAATTA	383
Q	y	312	caccatcataactcgcatggaanaaatctcgtgttggaagatatactctgaaactaogtga	371
D	b	384	CACCATATCAACTGCATGTGGAACAAATTCCTGGTTGGAAGATCTCGAATCTAGCGTGA	443
Q	y	372	aatcatctgacacttgtaggaacaactcaagaaatccaacagcggtgtgtatacaaatctc	431
D	b	444	AATATCTGAGCTCTGCAAGACACTAGGAAATCAACACGCTGTTGTTTCAAAATTA	503
Q	y	432	tcagaatgatacaacatctctgtactacatacaatcgttgatcttcgttaccatcaccaaca	491
D	b	504	TCAGATGATCAACATCTCTGATACATCAATCATGCTGATCTTGTTACATCACCACAA	563
Q	y	492	tcgtctgaaataactccaaatctacatcaacgycgctcgtgatacgcacgaacccgaattc	551
D	b	564	TCGCTCAATTAATCCCAAAATTCATCAACGCGCTCGATCGACCAAAACCGATCTC	623
Q	y	552	caatctggtgaacaacccaacgtccataaatacaatcgaatcgaacggaagttgtcgtga	611
D	b	624	CAATCTGGGTACATCCACGCTTCTAATAATCAATCAATGTTCAAACTGAGAGGTTGCGTGA	683
Q	y	612	cactcacgcgtacactctgtgatacaataacttcaactctgttcgcaaaagaactcgaacgaana	671
D	b	684	CACTCACCGCTACATCTGATCAAAATTACTCAATCTGTTGCAAAAGAACTGAACGAA	743
Q	y	672	agaataaagaacttgtagcaacaacagttccaattcttggtatccctggaagaacttctggg	731
D	b	744	AGAAATCAAAAGCCTGTGACCAACAACAGTCCAAATTCCTGATCTCTGAAAGACTTCTGGGG	803
Q	y	732	tgactacctcaagtagaacaacacgcgtactacaatgctgtagaactgtgacagatccgaacaaata	791
D	b	804	TGACTACTCTGACTGACGACAAACCGTATCTAATGCTGAAATCTGTACGATCCCAAAATA	863
Q	y	792	cgttcgactcaacaatctaggtatccgcggttcacatgtaaccctgaaaggtccgcggtgtc	851
D	b	864	CGTTGACGTCAACAACTTATAGTATCCGCGTTATCATGTACGTGAAGAGTCCCGGTGTTCC	923
Q	y	852	tgtaagtaaccaacaacttacctggaactcttcctctgatacsgtlyglaccnaatltcaat	911
D	b	924	TGTTATGACTACAAACATCTACCTGAAACTTCTCCGTACCGCTGGACCAATTAATCATAT	983
Q	y	912	caagaataacgctctcgtgtatacaagaacaatatcgttcgcaacaatgatacgtgtataat	971
D	b	984	CAAAATAATAGCGCGTCTGTGTAACAAGACAAATATCTGTTCCCAACAATGATCGTTATAT	1043
Q	y	972	caatgttgtatgaagaacaagaatacagctcgtgatacgaatgcttctcagctcgtggt	1031
D	b	1104	AGAAAGATCTTGTGCTCTCTGTGAAATCCCGGACGTTGTATATCTGCTCAAGTATGTT	1163
Q	y	1032	agaaaagatcttgctcgtctctggaatcccggaacgcttggtaatctgtctcagtagtgtc	1091
D	b	1164	AATTAATAATCCAAAGACACCGAGGTATCATCTAAATCAAAATGAATCTTCAGACGACA	1223
Q	y	1092	aatgaataccaagaagaccaaaggtatcacttaacaatatgcaaatgaaatctgcgagacaa	1151
D	b	1164	AATTAATAATCCAAAGACACCGAGGTATCATCTAAATCAAAATGAATCTTCAGACGACA	1223

```

QY 1152 caatgtaacgatcgttccgttccacacgtaacacatcgttaactgtc 1211
DB 1224 CATGTAACGATATGCGTTTCCATCGTTCCACGATCAACATATCCCTAACTG 1283
QY 1212 tgcctcaactggtlacaatcgttcgaacgctcctcgcactcgtgtgtcttg 1271
DB 1284 TGCTTCAACTGGTACAAATCGTCAGATCGAACGCTCTCTCCGACATCGGTGCTCTTG 1343
QY 1272 ggaagttcacccggttgatgacggttgagggtgaacgtcgcgtgtaa 1317
DB 1344 GGAGTTCAATCCGGTTGATGACGTTGGGGTGAACGTCCGCTGTAA 1389

RESULT
US-08-480-604A-27
; Sequence 27, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLTA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3888
; US-08-480-604A-27

Query Match          54.9%; Score 726.4; DB 1; Length 3891;
Best Local Similarity 72.3%; Pred. No. 1.3e-203;
Matches 943; Conservative 0; Mismatches 361; Indels 0; Gaps 0;

QY 14 tgcctcacttcacgtatcatcaagaacacatcacaactcctcgcactcgtcgt 73
DB 2588 TATCTACATTTACTGATATATATAGAAATTTATTAFACTCTATATGCAATTAAGAT 2647
QY 74 acgaatccaatcaaccgtacgactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 133
DB 2648 ATGAAAGTAATCATTTAAATGACCTTATGATGATCATCAAAATAAATATTTGGTAGTA 2707
QY 134 aagttacattcgatccgacgcgaagaatcgaacgtcgtcgtcgtcgtcgtcgtcgt 193
DB 2708 AAGTAAATTTTGATCCAAATGATGATAAATCAATTCATATTTATTTAGAAAGTAGTA 2767
QY 194 aaatcgaagttatccctgaagaatgctatcgtlatacaactatgtaacgaactcctca 253
DB 2768 AAATGAGGTAATTTTAAAAAATGCTATTTGTATATAGTAGTATGTAATAATTTTAGTA 2827
QY 254 cctccttcgtggtcgtatccgaaatcactcaccctcctcgtcgtcgtcgtcgtcgtcgt 313
DB 2828 CTAGCTTTTGGATTAAGAAATTCCTAGCTATTTTAACTATTAAGCTTAATAATTAAGATTA 2887
QY 314 ccaatcaactcgtcgtcgaagaacacatcgtgtggaagaatcctcgaactcgtcgtgaa 373
DB 2888 CAAATTAATTAATTTATGAGAAATAATTTAGAGATGGAAGATACCTTAATTAATGAGTAA 2947
QY 374 tcaatcgtgactcgtcgaacacacgaagaacacgaacgtgtgtcgtcgtcgtcgtcgt 433
DB 2948 TAATCTGAGACTTTTACAGGATCTCAGGAAATTAAGAAAGAGTAGTTTAAATTAACAGTC 3007
QY 434 agatgatacaactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 493
DB 3008 AAATGATTAATTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3067
QY 494 gtcgtataactcctcaaatcctacatcaacgcgcgtcgtcgtcgtcgtcgtcgtcgtcgt 553
DB 3068 GATTAAATATACCTCTAAATTTATATTAATGAAGATTAATGAATCAAAATCAATTTCAA 3127
QY 554 atctggttaacatccacgctccttaataacatcgtcgtcgtcgtcgtcgtcgtcgtcgt 613
DB 3128 ATTTAGCTAATTAATTCATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 3187
QY 614 ctcaacgcctacatcgtgatcaaatcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 673
DB 3188 CACATGATTAATTTTGGATTAATTAATTTTATTTTGTGTAAGATTAATTAATTAATTA 3247
QY 674 aaatcaagaacgtcgtcgaacacagtcgaactcgtcgtcgtcgtcgtcgtcgtcgtcgt 733
DB 3248 AAATCAAGATTTATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3307
QY 734 actaactcgtacgtcgaacacagtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 793
DB 3308 ATTAATTTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3367
QY 794 ttgaacgtcaacaatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 853
DB 3368 TCGATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3427
QY 854 ttatgactacacacatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 913
DB 3428 TAATGACTCAAAACATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3487
QY 914 agaaataacgctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 973

```


QY	674	aaatcaaaagccgttgaagaaaccagttccaattctgtgatactctcgaagaacttcgggtg	733
Db	3248	AAATCAAAAGATTATATATGATTAATCAATCAAAATTCAGATATTTTAAAGACTTTTGGGGTG	3307
QY	734	actaccgcagtaacgcaaaacccgttactaatactgtaatctgacatccgaacaatatcg	793
Db	3308	ATTATTTACAAFTATGATTAAMCCATCTATTATGTTAAATTTATATGATCCAAATAAATATG	3367
QY	794	ttgaagctcaacaatgatagatataccgcggttacaatgtacatcgaagaagtcgcggtttcgt	853
Db	3368	TGCGATGTAAATTAATGATAGGATTTAGAGGTTAATATGATCTTAAAGGGCTTAGAGGTACGC	3427
QY	854	ttatagcttacaacatcttaccgttaactcttcctctgacgcgttgatgccaatcatcatca	913
Db	3428	TATATGACTCAAAACATTTATTTAAATTCAGATTTGATATAGGGGGCAAAATTTATTAATA	3487
QY	914	agaaatacgcgtctgtglaaagaagacaataatcgttcgacaacaatgatcgtgtatataca	973
Db	3488	AAAATATTCCTCTGCAAAATAAAGATTAATATTTGTTAGAAATTAATGATCGTATATATTA	3547
QY	974	atgtgtgagttaagaacaagaatlaaccgctctggtctacaatgctctcaagctgtgtag	1033
Db	3548	ATGTAGTAGTTAAATAATAAAGAAATTAATAGGTTATAGCTAATATCATACACAGGACGGCTAG	3607
QY	1034	aaaagatctgtctgctcgtbgaatcccggaacgttgtaactctgtcgaagtatgttaa	1093
Db	3608	AAAAAATACTAAGTCTTATGAAATACCTGATGTAGGAATTAATGATCAATGATAGTAGTA	3667
QY	1094	tgaatccaagaagaccagggatcatcactaacaatgycanaatgaaatctgcagaaca	1153
Db	3668	TGAAGTCAAAAAATGATCAAGGATTAACAAATAATTAATGCAAAATGAAATTTACAAAGATATA	3727
QY	1154	atgttgacagatcaggtttcaatcaggtttccaacaagtccaataatcgtctaacaagtgtg	1213
Db	3728	ATGGGAATGATATAGGCTTTATATAGGATTTTATCATTGATTTAATATATATAGCTTAACATAGTAG	3787
QY	1214	cttccaacttggtaacaatcgtcagatcgaagaagttccctctcgaactctgggttgtctctgtg	1273
Db	3788	CAATTAATTTGTTAATATAGCAAAATAGAAAGATCTAGTAGCACTTTGGGTTGCTCATGGG	3847
QY	1274	agttcatcccggtgtgatgaagcgtttgggtgtgaacgctccgctgttaa	1317
Db	3848	AATTTATTTCTGTAGATGATGATGGGGGAAAAAGGCCACTGTATA	3891

RESULT 10
 US-07-618-312A-3
 Sequence 3, Application US/07618312A
 Patent No. 5389540
 GENERAL INFORMATION:
 APPLICANT: Makof Dr, Andrew J
 APPLICANT: Romanos Dr, Michael A
 APPLICANT: Clare Dr, Jeffrey J
 APPLICANT: Fairweather Dr, Neil F
 TITLE OF INVENTION: VACCINES
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: 14th Floor
 STREET: 2200 Clarendon Boulevard,
 CITY: Arlington,
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22201
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/618.312A
 FILING DATE: 19910516
 CLASSIFICATION: 424

```

1 PRIOR APPLICATION DATA:
2 APPLICATION NUMBER: GB 8926832.0
3 FILING DATE: 28-NOV-1989
4 PRIOR APPLICATION DATA:
5 APPLICATION NUMBER: GB 9006097.1
6 FILING DATE: 17-MAR-1990
7 ATTORNEY/AGENT INFORMATION:
8 NAME: Crawford Mr, Arthur R
9 REGISTRATION NUMBER: 25,327
10 REFERENCE/DOCKET NUMBER: 510-51
11 TELECOMMUNICATION INFORMATION:
12 TELEPHONE: 0101 703 8750400
13 TELEFAX: 0101 703 5253468
14 TELETYPE: 200797 NIXN UR
15 INFORMATION FOR SEQ ID NO: 3:
16 SEQUENCE CHARACTERISTICS:
17 LENGTH: 1359 base pairs
18 TYPE: NUCLEIC ACID
19 STRANDEDNESS: double
20 TOPOLOGY: linear
21 MOLECULE TYPE: cdna
22 HYPOTHEICAL: NO
23 ANTI-SENSE: NO
24 ORIGINAL SOURCE:
25 ORGANISM: Clostridium tetani
26 FEATURE:
27 NAME/KEY: CDS
28 LOCATION: 1..1356
29
30 US-07-618-312A-3

```

Query Match	12.6%;	Score 166.4;	DB 1;	Length 1359;
Best Local Similarity	52.3%;	Pred. No. 3.1e-39;		
Matches 486;	Conservative 0;	Mismatches 411;	Indels 33;	Gaps 4;

OY	40	aaatcatcaaatcaactccatccctgagaccggtcagcaatccaatcaactcgtgacgctg	99
Db	49	ATCCGAAAAAAAGTCAATCCATTCTTGAACTTTGGACATCAACACGATATTAATCTCCGACATC	108
OY	100	tctcgtcaactcgtcccaaatcaacatcgtgttctaaagthaactcgtalcgcg--atcgac	156
Db	109	TCTGGTTTCAACTCTCTTGTTATTCATATTCAGATGCTCAATTTGGTGGCCGGCATCAAC	168
OY	157	aagaatcagatccagctgttccaatcttgaatcttccaatatgaagttatccttgaagaat	216
Db	169	GGCAAGATATCCACCTGGTTAAACACGAAATCTTGTGAAGTTATGCGTGCACAAAGGCCATG	228
OY	217	gctatcgtatacaactctatcagaaaaactctccactccctctgtatcgtatcccg	276
Db	229	GACATCGAATACACGACATGTTCAACAACATTCCACCGTTAGCTTCTGGCTGGCGCTCCG	288
OY	277	aaatactcaactcctatc-----tctctgaacaatlgaaatacaacataccaactlgc	327
Db	289	AAAGTTCTCGCTCCACCTGGACAGTAAAGGACGACTTAACGAGTACTCCATCATCAAGCTCT	348
OY	328	atggaataacaat-----tctgtttgaaagtatctctgaactcaaggtgaa	372
Db	349	ATGAAGAAACATCCCTGTCCATCGGCGCTGGTTGGTGTCTGTTCCCTGGAAGGATAAACAC	408
OY	373	atcatcgtgaactctgaagacactcagggaataccaacagcgtgtgtlataccaatactct	432
Db	409	CTGATCTGGAGCTCTAAAGACACTCGCGGGGCAAGTTTCGTCATATCATCACTTCCGGCACCTG	468
OY	433	cagatgatcaacaactctcgtactactaatacgtcgtgatacttcgttataccaataccaat	492
Db	469	CCGGCAAGTTTCACGCGCTACTCTGGCTTAACAATAGGGTTTTCATCACTATCACTTAACGAT	528
OY	493	cgctcgaataactccaanaatactacataccaagcgctctgatactgacagaacccgctccc	552
Db	529	CGTCTGTTCTTCTGCTAAACCTGTACATCAACGGCGTTTCGATGGCGTCCGCTAAATCACT	588
OY	553	aatctggglaacataccagctctctlaataacatacatttcaaatcttgaacgtgtgtcgttgc	612

```

Db 589 GGTGTGGGGCTATTCGGTAGAGCAACAACATCCTTAAAGCTGAGACCGGTGCAACAC 648
QY 613 actcaccgtacatctgtagtcaaatctcaatctgttcgacaaagaactgaacgaataa 672
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AACAAACAGTACGTATTCATCGACAGATTCCTGATTTCTGCAACACACGAAACCGGAA 708
QY 673 gaaataaagacctgtagacaaacagtcacaaattctgtatctcgtgaagaacttcgggt 732
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 GAGATCGAAAACTGTATTCAGACGTACCTGTATTCACCTCTCGCGTACTCTGGGGGT 768
QY 733 gactaccctcagtagacaaacacgtagtactacatctgaatctgtagatccgaacaaatc 792
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AACCCGCTGCTTACGACACCGAATATTCCTGATCCCGGTACTCTTACTTAAAGAC 828
QY 793 gttagcgtcaacaatgtagtaccgcggtttacatgtacatgaagaagtcgcggtgttct 852
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 GTTACACTGAAAA-----CATCACTGACTACATGATGTACTGACCAACGGCGCTCTAC 882
QY 853 gttagtactacaacatctcaatctgaactcttccctgtaccgtgtgtaccaaattcattc 912
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 ACTAACGGTAACCTGAACATCTACTACCGACGTCTGTACMACCGCTGAAATTCATCATC 942
QY 913 aagaataacgcgtctgtgttaacaagacaat 942
Db 943 AACGCTACACTCCGAACACGAATTCGAT 972

```

```

RESULT 11
US-08-280-228-3
; Sequence 3, Application US/08280228
; Patent No. 5571694
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHIVE P.C.
; STREET: 1100 No. 5571694th Glebe Road
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,228
; FILING DATE: 25-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/618,312
; FILING DATE: 27-NOV-1990
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mary J
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELE: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 3:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-08-280-228-3

Query Match      12.6%; Score 166.4; DB 1; Length 1359;
Best Local Similarity 52.3%; Pred. No. 3, le-39;
Matches 486; Conservative 0; Mismatches 411; Indels 33; Gaps 4;

QY 40 aacatcatcaatacctccatccctgaacactgcgttagcgaatccaatcaccatgcagctg 99
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49 ATCCTGAAAAAGTCAACATTTGAACTTGACATCGATCAACAACGATATTATCTCCGACATC 108
QY 100 tctcgtagcgtcccaaaatcaacatccggttcttaagttaacttcgacg---atcgac 156
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 TCTGTTTCAACTCCTCTGTTATCAGATATTCAGATGCTCAATTGGTGGCGGCATCAAC 168
QY 157 aagaatcagatccagctggttcaatctggaatctctccaatcgaaatctatccgaagaat 216
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 GGCNAAGCTATCCACCTGGTTAACACGAAATCTTGAAGTTATTCGTGCACAAAGGCCATG 228
QY 217 gctatcgtatacaactctatagtagaagaacttccacactcctctcgtgacatccgtatccg 276
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 GACATGGAATACAAACGACATGTTCAACAACACTCACCGTTAGTCTTGCGGTCCGCTTCCG 288
QY 277 aaataactcaactccatc-----tctctgaacaatgaatacaccatcaactatgc 327
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 AAAGTTTCTGCTCCGACCTGGAACAGTACGGCAGTACAGTACAGTACATCATCAGCTCT 348
QY 328 atggaataacaat-----tctggttgaagaatctatctctggaactaagctgtaa 372
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 ATGAAGAAACACTCCCTGTCATCCGCTGTGGTTGGTCTGTTCCCTGAAAGGTAACAC 408
QY 373 atcatctgagactctcagagacactcaggaatacaaacagcgtgtgtatcaataactct 432
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 CTGATCTGGACCTCTGAAGAACTCCGGCGGAGATTGCTGAGATCACTTCCCGACACTG 468
QY 433 cagatgatcaaatctctgactacatcaatcgcgtgatatctcgtttaccatcaccacaat 492
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 CCGGACAAAGTTCAACCGGTACCTGGGCTTACMAAATGGGTTTTCATCACTATCACTACGAT 528
QY 493 cgtctgaataacctcaaaactatacaacgagcgtctgtagcaccgaacacgactctcc 552
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 CGTCTGTTCTTCTCTAACCCTGTAATCAACGGGCTTGTATGGGCTCCGCTGAATACAT 588
QY 553 aatctggttaacatccacgcttctaataacatcatgtttcaaacgtgacggtgtgtcgtac 612
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 GGTCTGGGCGCTATCCGTGAGACAAACAACATCACTTAAAGCTGGACCTTGCAACAC 648
QY 613 actcaccgtacatctgtagtcaaatctcaatctgttcgacaaagaactgaacgaataa 672
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AACAAACAGTACGTATTCATCGACAGATTCCTGATTTCTGCAACACACGAAACCGGAA 708
QY 673 gaaataaagacctgtagacaaacagtcacaaattctgtatctcgtgaagaacttcgggt 732
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 GAGATCGAAAACTGTATTCAGACGTACCTGTATTCACCTCTCGCGTACTCTGGGGGT 768
QY 733 gactaccctcagtagacaaacacgtagtactacatctgaatctgtagatccgaacaaatc 792
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AACCCGCTGCTTACGACACCGAATATTCCTGATCCCGGTACTCTTACTTAAAGAC 828
QY 793 gttagcgtcaacaatgtagtaccgcggtttacatgtacatgaagaagtcgcggtgttct 852

```

Db 829 GTTACGCTGAAAA-----CATCAGCTACATGTAACGACCAACGCCGCTCTAC 882
Qy 853 gtatgataccaacatctaccctgacccctgacccgtgtaaccaatcatc 912
Db 883 ACTAAGGTAACTGAACTATCTACTACGACGCTGTGTAACAGCGCTTAATTCATCATC 942
Qy 913 aagaatacgcctctgtaacaagaacat 942
Db 943 AAAGCTACACTCCGACACACGAATGCAT 972

RESULT 12
US-07-618-312A-1
; Sequence 1, Application US/07618312A
; Patent No. 5389540
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 14th Floor
; STREET: 2200 Clarendon Boulevard,
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/618, 312A
; FILING DATE: 19910516
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crawford Wt, Arthur R
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 510-51
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 0101 703 8750400
; TELEFAX: 0101 703 5253468
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-07-618-312A-1

Query Match 5.6%; Score 74.2; DB 1; Length 1359;
Best Local Similarity 47.38; Pctd. No. 3.9e-12;
Matches 223; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

Qy 340 tctgttggaagatctctctgacacacggtgaatcatctgacactcgcaggaacacacg 399
Db 376 TCTGTTGGAGAGCTATCTACTTAAGGTAATTAATTAATGACCTTTAAAGATTCCGG 435
Qy 400 gaaatcaaacagcgtgtgtatcaatctcagatgatcaaatctctgactacac 459
Db 436 GGAGAAGTTAGACAATACTTTAGGATTTACCTGATTAATTAATGCTTATTATAGCA 495
Qy 460 aatcgtgatactctgtaacatcaacacacacacacacacacacacacacacacac 519
Db 496 AATTAATGGGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 555
Qy 520 aacgcgcctgctgacacgaacacacacacacacacacacacacacacacacacacac 579
Db 556 AATGAGTACTTATGGAAGTGCAGAAATTAATGCTTTAGGAGCTTATTAAGAGATTAAT 615
Qy 580 aacatcatgttcaaacctgacacacacacacacacacacacacacacacacacacac 639
Db 616 AATATAACATTAATACTAGATGATGATTAATTAATTAATTAATTAATTAATTAATTA 675
Qy 640 tcaatcgttgcacaaagac 699
Db 676 TTTAGGATATTTTGCAGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 735
Qy 700 tccaatctgatactcgaagacacacacacacacacacacacacacacacacacacac 759
Db 736 TTAATCTATTAATCTTTTAAAGACATCTTGGGAACCTTTACGATATATACAGATTA 795
Qy 760 tacatcgtgaatctgtaac 810
Db 796 TATTTAATACAGTAGCTTCTAGTTCTAAGATGATTAATTAATTAATTAATTAATTA 846

RESULT 13
US-08-110-786A-7
; Sequence 7, Application US/08110786A
; Patent No. 5443966
; GENERAL INFORMATION:
; APPLICANT: MAKOFF, Andrew Joseph
; TITLE OF INVENTION: Expression of tetanus toxin fragment C.
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye P.C.
; STREET: 1100 No. 5443966th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110, 786A
; FILING DATE: 23-AUG-1993 1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/777,337
; FILING DATE: 29-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB90/00943
; FILING DATE: 20-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8914122.0
; FILING DATE: 20 June 1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary J. Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-134
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100
 TELE: 200797 NIXN UR
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1359 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA
 ORIGINAL SOURCE:
 ORGANISM: Clostridium tetani
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1359
 US-08-110-786A-7

Query Match 5.6%; Score 74.2; DB 1; Length 1359;
 Best Local Similarity 47.3%; Pred. No. 3.9e-12;
 Matches 223; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

```

OY 340 tcctgttggaagatctctgaactcaggtgaatcatctgacctcgcaggaactcag 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 376 TCTGTTGGAGTGATCACTTAAGTATTAATGACTTAATAAGATTCCGGC 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 400 gaaatcaaacagcgtgtgtatccaatactcagatgatacaactctgactacatc 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 436 GGAGAGATTAGCAATAATTAATCTTTAGGATTTACCTGATTAATTTAATGCTTATTAGCA 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 460 aatcgctgatacttgctccatccacaacatcgctgataactcgaatctcacc 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 496 AATTAATGGGTTTTTATTAATCTATTAATGATTAATCTTCTGCTAATTTGATATA 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 520 aacggccgtctgatcgaccagaacacatctccaatctggtgaactcgaactc 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 556 AATGGAGTACTTATGGAGATGAGAAATTAATGATTAATGATTAATGATTAAT 615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 580 aacatcatgttcaaatgacggtgtgtgacactcaccgctacatctgacatcaatc 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 616 AATTAATACCTTAATAGATGATTAATTAATTAATTAATTAATTAATTAATTA 675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 640 ttcaatctgtcgacaagaactgacgaagaagaagaactcgaactcgaactcag 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 676 TTTAGATATTTTTCGAAACATTAATTCGAAAGATTTGAAATTAATACACATTTAT 735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 700 tccaattctgtaacctggaagactctggtgtgactacatcgaactcgaactcag 759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 736 TTAATTAACCTTTTAAAGACTCTGCGGAAACCTTTAGATATGATTAAGATTA 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 760 tcaatgctgaatctgacatcgacaacaatacgttgacgtcaacaatgta 810
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 796 TATTTAATACAGTAGCTTCTAGTCTTAAGATGTTCAATTTGAAATATA 846
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 14
 US-08-280-228-1
 Sequence 1, Application US/08280228
 Patent No. 5571694
 GENERAL INFORMATION:
 APPLICANT: Makoff Dr, Andrew J
 APPLICANT: Romano Dr, Michael A
 APPLICANT: Clare Dr, Jeffrey J
 APPLICANT: Fairweather Dr, Neil F
 TITLE OF INVENTION: VACCINES
 NUMBER OF SEQUENCES: 13
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NIXON & VANDERHYE P C
 STREET: 1100 NO. 5571694th Glebe Road
 CITY: Arlington,
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22201-4714
 COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/280,228
 FILING DATE: 25-JUL-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/618,312
 FILING DATE: 27-NOV-1990
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 8926832.0
 FILING DATE: 28-NOV-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9006097.1
 FILING DATE: 17-MAR-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Wilson, Mary J.
 REGISTRATION NUMBER: 32,955
 REFERENCE/DOCKET NUMBER: 117-163
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 816-4000
 TELEFAX: (703) 816-4100
 TELE: 200797 NIXN UR
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1359 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Clostridium tetani
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1356
 US-08-280-228-1

Query Match 5.6%; Score 74.2; DB 1; Length 1359;
 Best Local Similarity 47.3%; Pred. No. 3.9e-12;
 Matches 223; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

```

OY 340 tcctgttggaagatctctgaactcaggtgaatcatctgacctcgcaggaactcag 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 376 TCTGTTGGAGTGATCACTTAAGTATTAATGACTTAATAAGATTCCGGC 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 400 gaaatcaaacagcgtgtgtatccaatactcagatgatacaactctgactacatc 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 436 GGAGAGATTAGCAATAATTAATCTTTAGGATTTACCTGATTAATTTAATGCTTATTAGCA 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 460 aatcgctgatacttgctccatccacaacatcgctgataactcgaatctcacc 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 496 AATTAATGGGTTTTTATTAATCTATTAATGATTAATCTTCTGCTAATTTGATATA 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 520 aacggccgtctgatcgaccagaacacatctccaatctggtgaactcgaactc 579
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 556 AATGGAGTACTTATGGAGATGAGAAATTAATGATTAATGATTAATGATTAAT 615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 580 aacatcatgttcaaatgacggtgtgtgacactcaccgctacatctgacatcaatc 639
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 616 AATTAATACCTTAATAGATGATTAATTAATTAATTAATTAATTAATTAATTA 675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 640 ttcaatctgtcgacaagaactgacgaagaagaagaactcgaactcgaactcag 699
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 676 TTTAGATATTTTTCGAAACATTAATTCGAAAGATTTGAAATTAATACACATTTAT 735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 700 tccaattctgtaacctggaagactctggtgtgactacatcgaactcgaactcag 759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 736 TTAATTAACCTTTTAAAGACTCTGCGGAAACCTTTAGATATGATTAAGATTA 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```


.....

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 21:30:08 ; Search time 7254.67 Seconds
(without alignments)
3824.928 Million cell updates/sec

Title: US-09-611-419a-5
Perfect score: 1326
Sequence: 1 gattcgaacagtgcctc.....aacgtcgcgtgaatc 1326

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_pa:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
------------	-------------	-------	--------	----	----	-------------

1	1302.2	98.2	1330	6	AR000029	AR000029 Sequence
2	1302.2	98.2	1330	6	AR169140	AR169140 Sequence
3	1302.2	98.2	1330	6	AX036243	AX036243 Sequence
4	1302.2	98.2	1338	12	XXU22962	022962 Synthetic b
5	1302.2	98.2	1402	6	AR000030	AR000030 Sequence
6	1302.2	98.2	1402	6	AR169141	AR169141 Sequence
7	1302.2	98.2	1402	6	AX036246	AX036246 Sequence
8	1302.2	98.2	1402	6	CB0707	CB0707 Sequence
9	727.2	54.8	4835	1	CB0707	CB0707 Sequence
10	726	54.8	3891	6	AR000031	AR000031 Sequence
11	726	54.8	3891	6	AR169142	AR169142 Sequence
12	726	54.8	3891	6	AX036248	AX036248 Sequence
13	661.6	49.9	4067	1	CBNT0X	CBNT0X Sequence
14	423.6	31.9	1299	12	AF251281	AF251281 Synthetic
15	422	31.8	1313	6	AS8946	AS8946 Sequence 6
16	227.8	17.2	4073	1	CB0707	CB0707 Sequence
17	227	17.1	1293	6	AS8945	AS8945 Sequence 5
18	227	17.1	4199	1	CB0707	CB0707 Sequence
19	227	17.1	4209	1	CB0707	CB0707 Sequence
20	194.8	14.7	3835	1	AB037704	AB037704 Sequence
21	194.8	14.7	3835	1	AB037705	AB037705 Sequence
22	194.8	14.7	3835	1	AB037706	AB037706 Sequence
23	194.8	14.7	3835	1	AB037707	AB037707 Sequence
24	194.8	14.7	3835	1	AB037708	AB037708 Sequence
25	194.8	14.7	3835	1	AB037709	AB037709 Sequence
26	194.8	14.7	3835	1	AB037710	AB037710 Sequence
27	194.8	14.7	3835	1	AB037711	AB037711 Sequence
28	194.8	14.7	3835	1	AB037712	AB037712 Sequence
29	194.8	14.7	3835	1	AB037713	AB037713 Sequence
30	194.8	14.7	3835	1	AB037714	AB037714 Sequence
31	194.8	14.7	4030	1	CBNT0X	CBNT0X Sequence
32	193.2	14.6	4017	6	AX088262	AX088262 Sequence
33	193.2	14.6	4017	6	AX088264	AX088264 Sequence
34	190	14.3	3838	1	CBNT0X	CBNT0X Sequence
35	190	14.3	3949	1	CBNT0X	CBNT0X Sequence
36	180.8	13.6	4234	1	CBNT0X	CBNT0X Sequence
37	180.8	13.6	9325	1	CBY13631	CBY13631 Sequence
38	166.4	12.5	1359	6	I28431	I28431 Sequence 3
39	166.4	12.5	3712	6	A49987	A49987 Sequence 4
40	166.4	12.5	3754	6	A37074	A37074 Sequence 17
41	166.4	12.5	3754	6	A42478	A42478 Sequence 17
42	166.4	12.5	3769	6	A37075	A37075 Sequence 18
43	166.4	12.5	4366	6	A49988	A49988 Sequence 5
44	166.4	12.5	4366	6	A42484	A42484 Sequence 12
45	166.4	12.5	4378	6	A42481	A42481 Sequence 9

ALIGNMENTS

RESULT 1
LOCUS AR000029 1330 bp DNA
DEFINITION Sequence 22 from patent US 5736139.
ACCESSION AR000029
VERSION AR000029.1 GI:3962560
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1330)
AUTHORS Kink,J.A., Thalley,B.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.
TITLE Treatment of Clostridium difficile induced disease
JOURNAL Patent: US 5736139-A 22 07-APR-1998;
FEATURES
source Location/Qualifiers
BASE COUNT 400 a 339 c 246 g 345 t
ORIGIN

Query Match 98.2%; Score 1302.2; DB 6; Length 1330;
Best Local Similarity 99.8%; Pred. No. 0;

Matches 1304: Conservative 0; Mismatches 3; Indels 0; Gaps 0;				
Qy	14	tgccctcactcctcactgatacatcaagaacacatcaatacctccatccctgaacctgc	73	
Db	11	tgctgcttacctcctcactgatacatcaagaacacatcaatacctccatccctgaacctgc	70	
Qy	74	gctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	133	
Db	71	gctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	130	
Qy	134	ctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	193	
Db	131	ctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	190	
Qy	194	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	253	
Db	191	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	250	
Qy	254	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	313	
Db	251	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	310	
Qy	314	acaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	373	
Db	311	acaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	370	
Qy	374	aaatcaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	433	
Db	371	aaatcaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	430	
Qy	434	ctcagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	493	
Db	431	ctcagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	490	
Qy	494	atcgaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	553	
Db	491	atcgaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	550	
Qy	554	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	613	
Db	551	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	610	
Qy	614	aaatcaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	673	
Db	611	aaatcaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	670	
Qy	674	aaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	733	
Db	671	aaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	730	
Qy	734	gtaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	793	
Db	731	gtaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	790	
Qy	794	agctgaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	853	
Db	791	agctgaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	850	
Qy	854	ctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	913	
Db	851	ctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	910	
Qy	914	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	973	
Db	911	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	970	
Qy	974	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1033	
Db	971	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1030	
Qy	1034	taagaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1093	
Db	1031	taagaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1090	

Qy	1094	taatgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1153	
Db	1091	taatgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1150	
Qy	1154	acaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1213	
Db	1151	acaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1210	
Qy	1214	tgctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1273	
Db	1211	tgctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1270	
Qy	1274	ggaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1330	
Db	1271	ggaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	1317	
RESULT 2				
ARI69140				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
FEATURES				
BASE COUNT				
ORIGIN				
Query Match				
Best Local Similarity 98.2%; Score 1302.2; DB 6; Length 1330;				
Matches 1304: Conservative 0; Mismatches 3; Indels 0; Gaps 0;				
Qy	14	tgccctcactcctcactgatacatcaagaacacatcaatacctccatccctgaacctgc	73	
Db	11	tgctgcttacctcctcactgatacatcaagaacacatcaatacctccatccctgaacctgc	70	
Qy	74	gctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	133	
Db	71	gctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	130	
Qy	134	ctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	193	
Db	131	ctgaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	190	
Qy	194	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	253	
Db	191	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	250	
Qy	254	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	313	
Db	251	ccaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	310	
Qy	314	acaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	373	
Db	311	acaagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	370	
Qy	374	aaatcaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	433	
Db	371	aaatcaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	430	
Qy	434	ctcagaatccaatccaatcctgatacctgctcctgctgaagctcccaaatcaaatcctgct	493	

QY 674 aagaatacaagaagctgtacagacaccagtcacattctgtatcctcgaagactctggg 733
|||||
Db 671 AAGAATAAAGAGACTGTACGACACACAGTCATATCTGGTATCTGAAAGACTTCTGGG 720
QY 734 gtagactactgacgtacgacaaacccgtacttaactgtctgaatctgtacgacgaacaat 793
|||||
Db 731 gtagactactgacgtacgacaaacccgtacttaactgtctgaatctgtacgacgaacaat 790
QY 794 aagctacgtacaaacatgtagtgcgcgcgtgtacatgtacacgtgaaggtccgcgtggt 853
|||||
Db 791 AGCTTACGTCACAAATGATAGATCCGGGTACATGATACCTGGAAGGTCGCCGTGGTT 850
QY 854 ctgtatgactacccaacatctacgtgaactcttcctgtacacgttggtaccacaatcaca 913
|||||
Db 851 CTGTTATGACTCTACCAACATCTACTGCTCCCTGTAACCTGGTACCAAAATTCATCA 910
QY 914 tcaagaataacgcgtctgttgaagaagaataatgcttcgcaacaatgtagtataca 973
|||||
Db 911 TCAGAAATACGCGTCTGTGTAACAGGACAAATATCGTTGCAACAAATGATCGTATACA 970
QY 974 tcaatgtgtatgtaagaacaaagaataacgcgtctgtacacgtcttcctcgaagctgtg 1033
|||||
Db 971 TCATGTGTGATGATAGAACAAAGATACCGTCTGGCTACCATGCTTCTCAGGCTGGTG 1030
Y 1034 tgaagaagaatctgtctgtctcgtcgtgaatacccggaagttgtgtaactgtctcgaagtgtg 1093
|||||
Db 1031 TAGAAAAGATCTTGTCTGTCTGGAATCCCGACCTTGGTAACTGTCTCAGGTAGTTG 1090
QY 1094 taatgaataccaaagaacgcgcgcgcgtgtacacgtacaaatgtaaatgtctcgaagaac 1153
|||||
Db 1091 TATATGAATCCAAAGACGACCAAGGATATCACTAACAAATGCAAAATGCAATCTCAGAGACA 1150
QY 1154 acaatgtgaacgatatcgttctcgtcgttcacacagttccacaatcatalcgtcctaactg 1213
|||||
Db 1151 ACAATGTGTAACGATATCGGTTTCATCGGTTTCACACAGTTCAACAAATGCTGTAACCTGG 1210
QY 1214 ttgcttccaaactggtacatcgttcaagatcgaagttcctcgtcgtcgtggtgtctct 1273
|||||
Db 1211 TTGCTTCCAACTGTGTAACATCGTCAATGTGAACGTTCTCTCCACTCGGTTGCTCTT 1270
QY 1274 ggaagttacatccgcgtgtgtagtcggtgtggaacgtcgcgtcgtgtaa 1320
|||||
Db 1271 GGGAGTTCATCCCGGTTGATGACGGTTGGGTTGAACGTCCGCTGTAA 1317

RESULT 4
XXU22962 1338 bp DNA linear SYN 08-NOV-1995
LOCUS
DEFINITION Synthetic botulinum neurotoxin serotype A Hc fragment (botA) gene,
complete cds.
ACCESSION U22962
VERSION U22962.1 GI:733428
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
synthetic construct.
artificial sequence.
1 (bases 1 to 1338)
REFERENCE
AUTHORS Clayton, M.A., Clayton, J.M., Brown, D.R. and Middlebrook, J.L.
TITLE Protective vaccination with a recombinant fragment of Clostridium
botulinum neurotoxin serotype A expressed from a synthetic gene in
Escherichia coli
JOURNAL Infect. Immun. 63 (7), 2738-2742 (1995)
MEDLINE 95310035
REFERENCE 2 (bases 1 to 1338)
AUTHORS Brown, D.R.
TITLE Direct Submmission
JOURNAL Submitted (17-MAR-1995) Douglas R. Brown, Toxicology, U. S. Army
Medical Research Institute of Infectious Diseases, Bldg. 1423, Fort
Detrick, Frederick, MD 21702-5011, USA
LOCATION/Qualifiers
1. .1338
/organism="synthetic construct"
/db_xref="taxon:32630"

/lab_host="Escherichia coli"
/note="based on Clostridium botulinum Type A neurotoxin
sequence"
9. .1325
/gene="botA"
9. .1325
/gene="botA"
/codon_start=1
/transl_table=11
/product="botulinum neurotoxin serotype A Hc fragment"
/protein_id="AAA80610.1"
/db_xref="GI:733429"
/translation="MARLSTFEYIKNIINTSLINLREYNHLLIDSLRYAKINIGS
KYNPDPIDKNQIOLEFNLEBSKLEVLKNAIVNSYENSTSFIRIPKYSISLNN
EYTIINCMEKNGMVKVSLNVEGILMTLOTOEIKORVYKROMINISDY IRRLEVT
ITNNLNNSKIYINGRLIDOKPILSINGNTHASNNIMFKIDGCRDHRHYIMVPLFD
KELNKEIKDLYDNDOSNGSLIDPFGNDYLOKPYMNLVDPNKVYVNNYNGJGGM
YIKGRGSMYNTNITLNSLYKGTFTIKTASGKNKDNIVRNNDRYIVNKKRBYR
LATVNSQAVEKILSALETPDVGNLSQVVVMSKNDQGITNCKMNLDDNNGNDIGFI
GPHQNNIATKIVASWYNNROIERSRSRTLCSSWEFTIPVDGNGERPL"

BASE COUNT 401 a 342 c 249 g 346 t
ORIGIN

Query Match 98.2%; Score 1302.2; DB 12; Length 1338;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 14 tggcctcctcctcactgatatcatcaagaacatcatcactcatcctgaactgc 73
|||
Db 19 TGCTGTCTACTCTCAGTATACATCAAGAACATCATCAATACCTCCTCAACCTGC 78
QY 74 gctacgaatcaatcacctcgtatcgcgtgtcgtcgcgttcacaaatcaaatcgtgt 133
|||||
Db 79 GCTACGAATCAATCAACTGATGACGCTGTGCTGCTACGCTTCCAAATCAATCAATCGGTT 138
QY 134 ctaaatctaactcgtatcgcgtatcgaagaatcaagatcgaactgttcaatcgtgaactct 193
|||||
Db 139 CTAAGATTAATCTCGATCCGATCGACAGAAATGATCCAGCTGTCAATCTGGAATCTT 198
QY 194 ccaaatcgaagtatccctgaagaatgcatgttatcaactatgtagaagaactctc 253
|||||
Db 199 CCAAAATCGAAGTTATCTTGAAGAAATGCTATACAACTTATGTAGCAAACTTCT 258
QY 254 ccaactcctcttgatccgtatcccgaaatcactcaactccactcctcgtgaacatgaaat 313
|||||
Db 259 CCACGTCCTCTGATCCGATCCGGAATATTCAACTCCATCTCTGTAACAAATGAAT 318
QY 314 acacacatcatcaactcgtatggaagaactcgtgtggaagaatgatatctcgtgaactcgtg 373
|||||
Db 319 ACACCATATCAATCATGCTGATGGAACAAATCTGTGGAAGATATCTGTAACCTGACGGTG 378
QY 374 aaatcatcgtgactcgtcgaaggaacactcagaagaatcagaagcgtgtgtatataact 433
|||||
Db 379 AAATCATCTGGACTCTGACAGGACCTCAGGAATTCAGACGCTGTGTTATTAATACT 438
QY 434 ctcaagttgtaacacatctcgtatcatcaatcgtcgtgactcgttctgtaacatcaacaaca 493
|||||
Db 439 CTCAGATGATCAACATCTGATCATCATGATGCTGATGCTGTTACATCACCACACA 498
QY 494 atcgtctgtaataactccaaaatctatcaacaagcgctgtgtagcgaacgaactc 553
|||||
Db 499 ATCGTCTGATTAATCTCAAAATCTATCATCAACGCGCTGTGATGACACGAAACCGATCT 558
QY 554 ccaactcgtgtgaatcaccacgcgtctcgaagaatcgaatcgaatcgaatcgaatcgaatcgaat 613
|||||
Db 559 CCATCTGGTATCATCCACGCTTCTTAATTAACATCAATGTTCAACATGAGAGGTTGTCGTG 618
QY 614 acactacgcgtacatcgtgatcaaatcactcgtcgtgtagcgaagaactggaagaa 673
|||||
Db 619 ACACGTCACGCTATCATCTGATCAATTAATCTGTTGTCGACAAAGAACATGGAAGAAA 678
QY 674 aagaatacaagaagctgtacgacacaccagtcacattctgtatcctcgaagaactcctggg 733

431 CTCAGATGATCAATCTCTGACTACATCAATCGCTGGATCTCTGTACCATCCACACA 490
494 atcgctgaataaaccacaatctacatacaacgcccgtctgtacccgaacacgatc 553
491 ATCGCTGAATTAAGTCCAAATCTCATATCAACGGCCGTGTGATCGACCAAGAACGATCT 550
554 ccaatctgggtaacatccagccttcaataacatcatgttcaaacatgagcgtgtgtcgtg 613
551 CCAATCTGGGTAAATCAACGCTTCTAATTAACATCATGTTCCTCAACCTGACGAGGTGGCTG 610
614 aaactccagcgtacatctcgatcaaatattcttctgtctgacaagaagcgtgaacgaa 673
611 AACCTACCGGTACATCTGGATCAATATCTTCAATCTGTCACAAAGACGTAACGAA 670
674 aagaatcaagaacctgtgacgaacaccagctcccaatctctgtatctcgtgaagaaccttgg 733
671 AAGAAATCAAGACCTGTACGACCAACAGCTCCAAATTCGTGTCTGAAAGACTTCTGGG 730
734 gtaactacctgtgacgaacaaacgttactatctgtgaatctgtacgaatccgaacaaat 793
731 GTGACTACTGCAATGACGAACAAACGCTACTACATGCTGTAATGTGATCGAATCGAACAAT 790
794 aggttgaagtcacaatgtaggatccgcggttacaatgtactctgaagaagtcgcgtgtgt 853
791 AGGTGACGTCAACATGTAGATATCCGGGTTACATGTACCTGAAAGGTCGCGGTG 850
854 ctgttatgactacaacatctaccgtgaacctctccctgtacggtgtgacgaatcaata 913
851 CTGTTATGACTACCAATCTACTGTAATCTTCCCTGTACCGTGTGATCCCAATTCATCA 910
914 tcaagaataacgctgtgtgtaacaaagacaatactgttcgacaacatgtatcgtatata 973
911 TCAGAAATACCGCTGTGTGTAACAAAGACAAATATCGTTGCAACATGATCGTATACA 970
974 tcaatgtgtagttaagaacaaagaataaccgtctgtgtactcaaatgtcttccagtcgtgtg 1033
971 TCAATGTTGTACTTAAGAAACAAAGATATCCGTGTACCAATGCTTCTGACGGCTG 1030
1034 tagaagaatctctgtctgtcgtgaacatccggaacgtgtgtatctgtctcagtgatgtg 1093
1031 TAGAAAAAGATCTTGTCTGTCTGTGGAATATCCGAGCTTGTGTATCTGTCTCAGTATG 1090
1094 taatgaatccaagaacgaacgggtatcatcaataacatgtcaaatgtcaatctgtcagga 1153
1091 TAATGAATCCAAGAACGACGAGGTATCACTAACAAATGCAATATGATTCGACAGACA 1150
1154 acaatgtgaagatagatcgttctcatcgtgttccacagttcaacaatatcgttaaatcgtg 1213
1151 ACAATGTGAACGATATCGGTTTCATCGGTTTCACCCAGTTCAACAATATCGGTTAAACTGG 1210
1214 tgcctcccaactgtgacaaatcgatcaatgaacgttccctcgtcgaactctgggttgcctt 1273
1211 TTGCTTCCACTGATGTAATCTGTCAATGTAACGTTCTCTCTGCACTCTGGGTTGCTCTT 1270
1274 gggagatctaccctgggtgtgacggttgggtgaacggttcggtcgtgtaa 1320
1271 GGGAGTTTCAATCCCGGTTGATGACGGTTGGGTAACGTCGCTGTAA 1317

RESULT 3
AX036243 1330 bp DNA linear PAT 16-NOV-2000
LOCUS AX036243
DEFINITION Sequence 22 from Patent EPI041149.
ACCESSION AX036243
VERSION AX036243.1 GI:11225862
KEYWORDS
SOURCE
synthetic construct.
ORGANISM
synthetic construct
artificial sequence.
REFERENCE
1 (bases 1 to 1330)
AUTHORS
Kink,J.A., Firca,J.R., Padhye,N.V., Thalley,B.S., Stafford,D.C. and
TITILE
Vaccine and antitoxin for treatment and prevention of C. Difficile

disease
Patient: EP 1041149-A 22 04-OCT-2000;
JOURNAL
OPHIDIAN PHARM INC (US)
FEATURES
source
1. 1330
/organism="synthetic construct"
/db_xref="taxon:32630"
1. 1317
/note="Synthetic"
/codon_start=1
/transl_table=11
/protein_id="CAC16478.1"
/db_xref="GI:11225863"
/translation="MARLSTFTEYIKNIIMTSILNLYESNHLIDLSRYASKINIGS
KYNFDPIDKNOIOLEFNESKIEVILKNAIYNSMYENSTFPIRIKPFNSISLN
EYTIINCENNSGKWSLVNGBEIIWTLDDOIEIKORVVKYSIMISDYINRIEYV
ITNRLNLSKIYINRLLDOKPISNLSGSHHSNNIMFLLDCGRHRYIKYFLEFD
KELNKEIKDLNDNSNSGILKDEWGDLYQDKPYMLNLTDPNRYVDVNVNGIEM
YLKGRGSVMPTNLIYNSLYRGTFIKKYSAGKNDIIVRNNDVYINVVYKNEYR
LATNASQAVEKIIISALEIPVGNLSQYVVKSKNDGKITKCKNLDNNGNDIGFI
GFHOENNTAKLVASNMVYRDIERSRTIGCSWEFIPVDGNGERPL"

BASE COUNT 400 a 339 c 246 g 345 t
ORIGIN

Query Match 98.2%; Score 1302.2; DB 6; Length 1330;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

14 tggcctctacacttactgaataatcaataagaacatcaataaccctccatccgtgacctgc 73
11 TCGTGTCTACTCTCACTGATATCAATCAACATCAATCAATCAATCAATCAATCAATCAAT 70
74 gctcgaatccaatccaccgtgacactgtctcgcacgcttccaaatccaatccatcgtgt 133
71 GCTACGAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 130
134 cttaagtlactctgactccgatccgaagaacatccagctgttcaatctggaactct 193
131 CTAAATTTAACTTCATCCATCGATCGACAGATCCAGATCCAGATCCAGATCCAGAT 190
194 ccaaaatcgaagttatcctcgtgaagaatgcatcgtatatacaactctatgtacgaactct 253
191 CCAAAATCGAAGTTATCTCGAAGAAATGTTATGTAACACTATGTAACACTATGTAAC 250
254 ccaactcccttctggtatccgtatccggaatctcaactccatctctcgaacaatgaat 313
251 CCACCTCTTCTGATCCGTAATCCGAAATFACTCACTCATCTCTGAAACATGAAT 310
314 acaacatcaactgaactgtgacgaacaaatctgtgtggaagatcctcgtgaactacgtgt 373
311 ACACCATATCAACTGATCGAAGAAACAAATCTGTGGAAGATCTGGAATCTGGAATCG 370
374 aaatcatctggaactctcgaagaactcgaagaatcaacaacagcgtgtgtatcaataact 433
371 AAATCATCTGGACTCTGACGACACTCAGGAATCAACAGCGTGTGTATTCATTAATCT 430
434 ctccagatgaataatctctgactacataatcgcgtggtcttgcgttaccatcaccaaca 493
431 CTCAGATGATCAACATCTGTGACTATCAATCGGTGATCTTCTGTACCATCAACACA 490
494 atcgctgaataaaccacaatctacatacaacgcccgtctgtacccgaacacgatc 553
491 ATCGCTGAATTAAGTCCAAATCTCATATCAACGGCCGTGTGATCGACCAAGAACGATCT 550
554 ccaatctgggtaacatccagccttcaataacatcatgttcaaacatgagcgtgtgtcgtg 613
551 CCAATCTGGGTAAATCAACGCTTCTAATTAACATCATGTTCCTCAACCTGACGAGGTGG 610
614 aaactccagcgtacatctcgatcaaatattcttctgtctgacaagaagcgtgaacgaa 673
611 AACCTACCGGTACATCTGGATCAATATCTTCAATCTGTCACAAAGACGTAACGAA 670

```

Db 679 AAGAAATCAAGACCTGTACGACAAACCACTCCAAATCTGTAATCCGAAAGACTTGGG 738
Qy 734 gtagactactgtagtagcgaacaacgtagtactatgctgtaactctgtacggtccgaagaat 793
Db 739 GTGACTACCTGGAGTACGCAAAACCGTACTACATGCTGAATCTGTACGATCGGAACAAAT 798
Qy 794 acgttaacgtaacaatgtagtattccggcggttactgtaactgtaagaaggtccgctggtt 853
Db 799 ACCTTGACCTCAACAATGTAGTATCCGGGTTTACATGTACTGAAAGGTCCGGCTGGTT 858
Qy 854 ctgttaatgactaaccaacatctacccctgaactctccctgtagcgtgtgtgtaacaattcata 913
Db 859 CTGTATGACTACCAACAATCTACCTGTAAGTCTCCCTGTACCGGTGTACCAATTCATCA 918
Qy 914 tcaagaataacgctgtctgttaacaagaagaataatgcttgcgaacaatgtagtataca 973
Db 919 TCAAGAAATACCGCTGTGTGTACCAAGGACAAATGCTTGGCAACAATGATCGTGTATACA 978
Qy 974 tcaatgttagttaagaacaagaataacacgctgtgtagtaacaaatgcttccagcgctggtg 1033
Db 979 TCAATGTTTACTTAAAGAACAAAGAAATACCGTGTGCTACCAATGCTTTCAGGCTGGTG 1038
Qy 1034 tagaaaagaatcttctgtctgtctgtgaatcccggaacgttgttaactctgttccagtagt 1093
Db 1039 TAGAAAAGATCTTGTCTGCTGTGAATCCCGAGCTGTGTATCTGTCTCAGGTAGTTG 1098
Qy 1094 taatgaataccaagaagcagcggtatataactaacaataatgcaaatgtaactgtcgaagaca 1153
Db 1099 TAATGAATAATCCAAAGAACGACGAGGTATCCATACAAATGAATGAATCTGCAGAGACA 1158
Qy 1154 acaatgtagaagatcgtttccatcggtttccacagttcaacaatatcgcttaacatcg 1213
Db 1159 ACAATGTGTACATATCGGTTTTCATCGGTTTCACACAGTTTCAACAATATCGCTAAACTGG 1218
Qy 1214 ttgcttccaactgtgtacaaatcgtlccagatcgaaacgttccctcgcactcctggttgcctt 1273
Db 1219 TTGCTTCCAACTGTAGACATCGTCAAGATGAAAGTTCCTCGCACTCTGGGTTGCTCTT 1278
Qy 1274 gggatattatcccggtgtgtgtaacggttgggtggaacgtccgctgttaa 1320
Db 1279 GGGATTCATCCCGGTGTGTGTGACGGTTGGGTTGAACGTCCGCTGTAA 1325

RESULT 5
AR000030 1402 bp DNA linear PAT 04-DEC-1998
DEFINITION Sequence 25 from patent US 5736139.
ACCESSION AR000030
VERSION AR000030.1 GI:3962561
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE
AUTHORS 1 (bases 1 to 1402)
TITLE Kink,J.A., Thalley,J.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.
JOURNAL Treatment of Clostridium difficile induced disease
FEATURES
location/Qualifiers
source 1..1402
BASE COUNT 420 a 360 c 260 g 362 t
ORIGIN

```

```

Query Match 98.2% Score 1302.2; DB 6; Length 1402;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy 14 tggctctactactatcatatcaagaacaataatcctccatcctgtaactgctg 73
Db 83 TGGCTCTACCTTCACTGATCATCAAGACATCATCAATACCTCCATCTGAACTTGC 142

```

```

Qy 74 gctacgaatccaatccactgtagactgtctcgtacgcttccaaatatcaatcgtgt 133
Db 143 GCTACGAATCCAAATCAACCTGATCGACCTGTCTCGTACGCTTCCAAATCAACATCGGTT 202
Qy 134 ctaagttaacttgcgtccgatcgcagaagaatcgaatccagcgtgttcaactctggaactt 193
Db 203 CTAAAGTTAACTTCGATCCGATCGACAAAGATCCAGCTGTTCAATCTGGAATCTT 262
Qy 194 ccaaaatcgaaatattccctggaagaatgtagtatacgaactctatgtcgaaacttct 253
Db 263 CCAAAATCGAAGTTATCTGGAAGAAATGTATATACAACTTATGTACGAAACTTCT 322
Qy 254 ccaactcctcttgatccgtagatccggaataactcaaccctctcctcgaacaatgaa 313
Db 323 CCACCTCTTCTTGATCCGATTCGCCAATTAATTACTTCACTCCATCTCTTGAAACATGAAT 382
Qy 314 acaacatcatcaactgcatgtgaaanaaacttcgtgtggaaglatcctgaaactacgtgtg 373
Db 383 ACACCATCATCAACTGCATGAAACAAATTCGTGTGAAAGATCTGAACTACGAGTG 442
Qy 374 aaatcatctggaactctggaagacactcaagaaatacaacaggtgtgtatccaatct 433
Db 443 AAATCATCTGGACTGTGACAGACACTCAGGAATCAACACGCTGTGTATTCAAATACT 502
Qy 434 ctcaagaatgaatcaactctgtagatcatcatcgtcgtgttctgttaccatcaaca 493
Db 503 CTCAGATGATCAACATCTCTGACTACATCATATCGCTGGAATCTTCTGTACATCACCACA 562
Qy 494 atcgtctgaataactccaataatcatatacaacggtccgtgtgatcgaacgaacgatct 553
Db 563 ATCGTGTGAATTAATCCAAATCTACATCAACGGCGCTGTGATCGACAGAAACCATCT 622
Qy 554 ccaatctggtgtaacatcaacagcttcttaataataatgttcaaacgtgagcgtgtgtcgtg 613
Db 623 CCAATCTGGGTAAACATCCACGCTTCTTAATTAATCAATCAATGTTCAACATGGACGCTGTCGTG 682
Qy 614 acaactcagcgtacatctgtatcaaatctcaatctgttctgcgaagaacgtgaagaa 673
Db 683 ACACCTACCGCTACATCTGTGATCAATTAATTCTGATCTGTTCGACAAAGAACTGAAGAAA 742
Qy 674 aagaaatcaagaactgttagcgaacaacagttccaattctgttatcctgaaagaactctggg 733
Db 743 AAGAAATCAAGACCTGTGTACGACAAACCACTCCAAATCTGTAATCTCGAAAGACTTGGG 802
Qy 734 gtagactactgtagtagcgaacaacgtagtactatgttgaatctgtacggtccgaagaat 793
Db 803 GTGACTACCTGGAGTACGCAAAACCGTACTACATGCTGAATCTGTACGATCCGAACAAAT 862
Qy 794 acgttgaacgtaacaatgtagtagatccggcggttactatgtaactgtaagaaggtccgctggtt 853
Db 863 ACCTTGACCTCAACAATGTAGTATCCGGGTTTACATGTACTGAAAGGTCCCGCTGGTT 922
Qy 854 ctgttaatgactaaccaacatctacccctgaactctccctgtagcgtgtgtgtacaaattcata 913
Db 923 CTGTATGACTACCAACAATCTACCTGAACCTTCCCTGTACCGGTGTACCAAAATTCATCA 982
Qy 914 tcaagaataacgctgtgtgttaacaagaagaataatgcttgcgaacaatgtagtataca 973
Db 983 TCAAGAAATACCGCTGTGTGTACCAAGGACAAATGCTTGTGCAACAATGATCGTGTATACA 1042
Qy 974 tcaatgttagttaagaacaagaataacacgctgtgtagtaacaaatgcttccagcgctggtg 1033
Db 1043 TCAATGTTTACTTAAAGAACAAAGAAATACCGTGTGCTACCAATGCTTTCAGGCTGGTG 1102
Qy 1034 tagaaaagaatcttctgtctgtgtgaatcccggaacgttgttaactctgttccagtagt 1093
Db 1103 TAGAAAAGATCTTGTCTGCTGTGAATCCCGAGCTGTGTATCTGTCTCAGGTAGTTG 1162
Qy 1094 taatgaataccaagaagcagcggtatataactaacaataatgcaaatgtaactgtcgaagaca 1153
Db 1163 TAATGAATAATCCAAAGAACGACGAGGTATCACTAACAAATGAATGTGCAGAGACA 1222
Qy 1154 acaatgtagaagatcgtttccatcggtttccacagttcaacaatatcgttaaacgtg 1213

```

|||||
Db 1223 ACAATGGTAACGATATCGGTTTCATCGTTCCACCAAGTTCAACATATCGTAAACTGG 1282
Qy 1214 ttgcttccaaactggtacaaatcgctcagatcggaagcttcctctgcatctgggttgcctt 1273
Db 1283 TTGCTTCCAACTGGTACATCGTCAAGATCGAAGGTTCTCTCGCACTCGGGTGTCTT 1342
Qy 1274 gggagttcaatcccggttgatgaagcttgggggtgaagctccgctgtaa 1320
Db 1343 GGGAGTTCAATCCCGGTTGATGACGGTTGGGGTGAACGTCGCTGTAA 1389

RESULT 6
AR169141
LOCUS AR169141 1402 bp DNA linear PAT 17-DEC-2001
DEFINITION Sequence 25 from patent US 6290960.
ACCESSION AR169141
VERSION AR169141.1 GI:17906911
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1402)
AUTHORS Kink, J. A., Thalley, B. S. and Stafford, D. C.
TITLE Vaccine and antitoxin for the treatment of C. difficile disease
JOURNAL Patent: US 6290960-A 25 18-SEP-2001;
FEATURES
source 1.1402
BASE COUNT 420 a 360 c 260 g 362 t
ORIGIN

Query Match 98.2%; Score 1302.2; DB 6; Length 1402;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 14 ttgcttctactctcaatgaatacatcaagaacatcatcaatcctccatctcgaacctgc 73
Db 83 TGTGTCCTACTCTTCACTGATCATCAAGAACATCATCACTCCATCTCGAAGCTGC 142
Qy 74 gctcgaatccaaatcaactcgaatcgatcgctcgaatcctcgaatcctcgaatcgaatcggt 133
Db 143 GCTACGATCCAAATCAACCGATGAGCCTGCTGCTACGCTTCCAAATCAATCAATCGGT 202
Qy 134 ctaaaatcaactcgaatcgaatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgat 193
Db 203 CTAAAGTTAACTCGATCCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 262
Qy 194 ccaaaatcgaaatcctcgaatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgat 253
Db 263 CCAAAATCGAAATTCGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 322
Qy 254 ccaactcctctcgtgacccgcatccgaaatcaactcgaatcgaatcgaatcgaatcgaat 313
Db 323 CCACCTCTCTTCTGGATCCGATCCGAAATCACTCAATCTCTGTAACAAATGAAT 382
Qy 314 acacatcatcaactcgaatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgatcgat 373
Db 383 ACACCATCAATCAATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 442
Qy 374 aaatcatctggaatcgtcgaagacatcgaaatcaaaacagcgtgtgtatcaaatcact 433
Db 443 AAATCATCTGAGCTCGAGGACATCAGGAATCAAAACAGCTGTGTATCAAAATACT 502
Qy 434 ctaagaatgaatcaatcgtcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 493
Db 503 CTGAGATGATCAATCTCTGATCAATCAATCTCTGATCAATCTCTGATCAATCTCTGATCA 562
Qy 494 atcgtctgaatcaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 553
Db 563 ATGCTGTAATCAATCTCAAAATCTACATCAACGCGCTGTGATCGACGAGAAACCGATCT 622

Qy 554 ccaatctgggttaacatccacgcctctcaataacatcattcgaatcgaatcgaatcgaatcga 613
Db 623 CCAATCTGGGTAACTACATCCAGGTTTAAATCAATCAATCAATCAATCAATCAATCAATCAAT 682
Qy 614 acacatccggtcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 673
Db 683 ACACCTACCGCTCAATCTGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 742
Qy 674 aagaatcaaaagacccgtgaagacacacgctcgaatcgaatcgaatcgaatcgaatcgaatcga 733
Db 743 AAGAAATCAAAAGACCTGTACGACACGATCAATCTGATCAATCTGATCAATCTGATCAATCT 802
Qy 734 gtaactcctcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 793
Db 803 GTGACTACCTCGCATGACGACAAACCGTACATCAATCTGATCAATCTGATCAATCTGATCA 862
Qy 794 agcttgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 853
Db 863 AGCTTGAACGTCAAAATGATGATCGGCTGATCAATCTGATCAATCTGATCAATCTGATCA 922
Qy 854 ctgtatgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 913
Db 923 CTGTTATGACTACCAATCAATCTGATCAATCTGATCAATCTGATCAATCTGATCAATCTGA 982
Qy 914 tcaagaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 973
Db 983 TCAAGAAATCGGCTGATGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1042
Qy 974 tcaatcttgatgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 1033
Db 1043 TCAATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1102
Qy 1034 tagaagaatcgtctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1093
Db 1103 TAGAAAAAGATCTTGTCTGCTGGAATCCCGACGCTGATGATGATGATGATGATGATGATG 1162
Qy 1094 taatgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcga 1153
Db 1163 TAATGAATTCGAAGAACGACCAAGGATCACTCAAAATGATCAAAATGATCAAAATGATCA 1222
Qy 1154 acaatgtaacgatacgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1213
Db 1223 ACAATGGTAACGATATCGGTTTCATCGTTCCACCAAGTTCAACATATCGTAAACTGG 1282
Qy 1214 ttgcttccaaactggtacaaatcgctcagatcggaagcttcctctgcatctgggttgcctt 1273
Db 1283 TTGCTTCCAACTGGTACATCGTCAAGATCGAAGGTTCTCTCGCACTCGGGTGTCTT 1342
Qy 1274 gggagttcaatcccggttgatgaagcttgggggtgaagctccgctgtaa 1320
Db 1343 GGGAGTTCAATCCCGGTTGATGACGGTTGGGGTGAACGTCGCTGTAA 1389

RESULT 7
AX036246
LOCUS AX036246 1402 bp DNA linear PAT 16-NOV-2000
DEFINITION Sequence 25 from Patent EP1041149.
ACCESSION AX036246
VERSION AX036246.1 GI:11225864
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1402)
AUTHORS Kink, J. A., Firca, J. R., Padhye, N. V., Thalley, B. S., Stafford, D. C. and Williams, J. A.
TITLE Vaccine and antitoxin for treatment and prevention of C. Difficile disease
JOURNAL Patent: EP 1041149-A 25 04-OCT-2000;
FEATURES
source 1.1402
/organism="synthetic construct"

CDS
/db_xref="taxon:32630"
1..1389
/note="Synthetic"
/codon_start=1
/transl_table=1
/protein_id="CAC16479.1"
/db_xref="GI:11225865"
/translation="MGHHHHHHSSGHIEGRHMASMARLSTPEYIKNIINTSI
LNLRYESNHLIDLSRVASKINISKYNPDDPKNOLOLFNESSKIEYILKNATVYNS
MYENFSTFWIRIPKRYNSISLNEYITINOMENNSGKWSLNGELIWTLODIELK
QRYVFKTSQMINISDIYINRWLEFVITNNRNNSKIYINGRLIDKPISNLGNHASNN
IMFKLIDGCRDTHRYIWKYFNLPDEKLENEKIKDLYDNOSNGSLDKDFWGDLYDYP
YMLNLDPNKYVYNNVNGIYKGYMYLKGPRCSVMYTNLYNSLSLRYGKFLIKRYASG
KNDIVRNNDREVYINVVYKNKEYRLATNASQAGVEKILSALEIFDVGLSOVVYMKSK
NDOGITKKCKMNIODNNGNDIGFISGFHFNFIATLVASNWNRIENSRILGCSMER
IPVDDGWERPL"

BASE COUNT 420 a 360 c 260 g 362 t
ORIGIN

Query Match 98.2% Score 1302.2; DB 6; Length 1402;
Best Local Similarity 99.8% Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 14 tggcctcactcactcactaatacaagaacatcatcaatcactccactcgaacctgc 73
DB 83 TGGCTGTACTCTCAGTAATACATCAAGAACATCATCATCTCCATCTGAACTGCG 142
QY 74 gctacgaatccaatcactgatcgactgtctcgctacgcttccaataaactcggtt 133
DB 143 GCTACGAATCCAAATCCTGATCGACCTGTCTCGCTACGCTCCAAAATCAATCGGTT 202
QY 134 ctaagtaactcactcactcactcgaagaacatcagaatcgaatcgaatcgaatc 193
DB 203 CTAAGTAATCTTCATCTCGATCGATCGACAGAAATCAGATCAGATCTGCAATCTT 262
QY 194 ccaaatcgaagtataccctgaagaatgtatcgtatacaactctatgtacgaataact 253
DB 263 CCAAAATCGAAGTATTCCTGAAGATGCTATCGTATACACGCTATGTAAGCAAACTCT 322
QY 254 ccaactcctcttgatccggtatcccgaaataactccaactcctctcgaataatga 313
DB 323 CCACCTCTTGTGATCGGTATCCCGAAATACTTCACCTCCATCTCTGAAATGTAAT 382
QY 314 aacacatcactcactcactcactcactcactcactcactcactcactcactcactc 373
DB 383 ACACCATATCACTGCTGATGGAACAAATCTGTTGGAATATCTCTGAACTACGGTG 442
QY 374 aaatcactcactcactcactcactcactcactcactcactcactcactcactcact 433
DB 443 AAATCATCTGACCTGCGAGGACACTCAGAAATCAAAACAGCGTGTGATTCAAATACT 502
QY 434 ctacagatgatacaactcctcactcactcactcactcactcactcactcactcactc 493
DB 503 CTCAATATCAATCAATCTCTGATCAATCAATCGCTGATCTCTTACCATCACACACA 562
QY 494 atcgtcgtataactcactcactcactcactcactcactcactcactcactcactc 553
DB 563 ATCGCTCAATTAATCTCAAAATCTACATCAACGCGCTGTGATCCAGAAACCGATCT 622
QY 554 ccaatcgtgatacaactcactcactcactcactcactcactcactcactcactcactc 613
DB 623 CCAATCTGGTAAATCAACGCGCTTCTAATACATCATGTTCAAACTGAGCGTGTGCGTG 682
QY 614 acactcaccgctacactcactcactcactcactcactcactcactcactcactcactc 673
DB 683 ACACCTACCGCTACATCTGATCAATCTCTTCAATCTCTTCCACAAAACGAAACGAAA 742
QY 674 aagaatacaagaccctgacgaacaacacgactcgaatctcgtatcctcgaagaactctgg 733
DB 743 AAGAATAAAGACCTGTACGACAAACAGTCCAAATCTCTGATCTCTGAAAGACTCTTG 802
QY 734 gtgactacctgacgtacgaacaacccgtactacatgctgaatctgtacgatccgaacaat 793

DB 803 GTGACTACTCGTAGTACGAAACCGGTAATGATGTGATGATCGGAAACAAAT 862
QY 794 acgttgaacgtacaactatgaggtatccgcggttacaatgataaagtcgcgtgtt 853
DB 863 AGCTTGACCTCAACATGATGATGATCCCGGTTTACATGATCTGAAAGGCTCCGCGTGGT 922
QY 854 cgttatgactcacaactcactcactcactcactcactcactcactcactcactcactc 913
DB 923 CTGTATGACTCAACATCTACCTGAACTCTTCCCTGACCGTGGTACCAATTCATCA 982
QY 914 tcaagaataacacgcgtcgttgaacaagaacataatcgttcgacaacatgacgtgtataca 973
DB 983 TCAAGAAATACCGCGTGTGTATCAAGAACAAATATGTTGCGAACATGATCGTATATCA 1042
QY 974 tcaatggttagttaaagaacaagaataacacgctcgtgactcaaatgcttccacgctgtgt 1033
DB 1043 TCAATGTTGTACTGTAAAGAAACAAAGAAATACCGTGTGCTACCAATGCTTCTCAGCGTGGTG 1102
QY 1034 tagaagaagacttgcctcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1093
DB 1103 TAGAAGAAGATCTTGTGTCTGTAATCCCGGAGTGTGTATCTGTCTCAGTAGTTG 1162
QY 1094 taatgaatccaagaacacgacggtatcactaacaataatgcaaaatgaaatcgtcagagaca 1153
DB 1163 TAATGAATCCAAGAACGACGAGGATCATCAACAAATGCAAAATGAAATCTGACAGACA 1222
QY 1154 acaatggtgaagatatacgttccatcactcgttccacacggttcaacaataatcgtcgaat 1213
DB 1223 ACAATGTTAAGCATATCGGTTTCAATCGGTTTCCACAGTTCAACAAATATCGCTAAACATG 1282
QY 1214 ttgcttccacacggtgatacgttcacatcgaatcgaatcgttccacacggttccatcgtt 1273
DB 1283 TTGCTTCCACACTGTTACAAATCTGATGAAACGTTTCTCTCTGCACTCTGGGTTGCTCTT 1342
QY 1274 gggagttcaatcccggttgaacggttgggtgaaatcgtcgtgttaa 1320
DB 1343 GGGAGTTCAATCCCGGTTGATGACGGTTGGGGTGAACGTCGCTGTAA 1389

RESULT 8
CBROTAG 4292 bp DNA linear BCT 12-SEP-1993
DEFINITION Clostridium botulinum bota gene for type A neurotoxin.
ACCESSION X52066.1 X52088
VERSION X52066.1 GI:40381
KEYWORDS bota gene; neurotoxin; secreted protein.
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
REFERENCE 1 (bases 1 to 4292)
AUTHORS Minton,N.P.
TITLE Direct Submission
JOURNAL Submitted (08-JAN-1990) Minton N.P., PHLS Centre for Applied Microbiology & Research, Molecular Genetics Group, Division of Biotechnology, Porton Down, Salisbury SP4 0UG Wiltshire, U K
REFERENCE 2 (bases 1 to 4292)
AUTHORS Thompson,D.E., Brehm,J.K., Oulttram,J.D., Swinfield,T.J.,
Shone,C.C., Atkinson,T., Mellings,J. and Minton,N.P.
TITLE The complete amino acid sequence of the Clostridium botulinum type A neurotoxin, deduced by nucleotide sequence analysis of the encoding gene
JOURNAL Eur. J. Biochem. 189 (1), 73-81 (1990)
FEATURES
MEDLINE 90235864
FEATURES location/Qualifiers
1..4292
/organism="Clostridium botulinum"
/strain="sub sp. type A, NCTC2916"
/db_xref="taxon:1491"
16..64
misc_feature
promoter 19..24

ORGANISM Clostridium botulinum
 Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 Clostridium.
 REFERENCE 1 (bases 1 to 4835)
 AUTHORS Blaz,T., Kurazono,H., Wille,M., Frevert,J., Wernars,K. and
 Niemann,H.
 TITLE The complete sequence of the botulinum type A neurotoxin and its
 comparison with other Clostridial neurotoxins
 JOURNAL J Biol. Chem. 265, 9153-9158 (1990)
 COMMENT 90264400
 FEATURES
 source
 Location/Qualifiers
 1..4835
 /organism="Clostridium botulinum"
 /db_xref="taxon:1491"
 240..>4835
 /product="neurotoxin mRNA"
 344..349
 /note="ribosome binding site"
 358..4248
 /note="neurotoxin"
 /codon_start=1
 /transl_table=1
 /protein_id="AA03262.1"
 /db_xref="GI:144865"
 /translation="MPFYVKQFNKDPYNGVDIAYIKIRPNAGOMQPVAKRIHKIKW
 IPEBDFTNPEEDLNPREAKQVPVSIDSTYSTDNKKNYLKGVKLEIRITSD
 LGRMLTIVRGIFPWGSGTIDTELKVIDICINIQDQSYRSELMVLIGPADPA
 IOECKSGFGEVNLTRNGYSTQYIRSPFTGFESELEVDNPLIGAKFADPA
 VTLHELHAGHRLYGIAINPVRKVTNAYEMSGLEVFEELRTFGDGAFTDS
 LOENEPFLYYNKKFKDIASLNKASKYGTASIOALNVEKKEKLSLEDGKFSV
 KIKEDKLYKMLLETEDNEVYKFKVLNKRKYELNDAKVEFINIPKPNYIYQSEN
 RNTMLAANGQNEINMNTKLNKFTGEPKILDCRITISKRTSLDKYKAL
 NDLCIKVNMDFSPEDNFTDLNKEEITSPNIEAEENSLDILQYLYLTF
 DNEPENSISLSSDILIGLELMPNIEFKPKGKYEIDKVTMFHYLRQDEHKSRI
 ALTNSVNEALNPSRVYTFSSDYVKVKNKPKTEAMFGVQEQIAYDFTDSEVST
 DKIDITIIIPYIGPALNIGMLKDEPGALITSGAVILEFPEIPIVIGTFALV
 SYLNKVLVQTIQNALSKREKMDQVYKIVTMNAVNTQIDILRKMKAELENOA
 EATKAITVOYQTEEEKNNINIDLSKINSEIKAMINIKPLNOSYSILAN
 SMIPYKVRLEDYFASLKDALKITTYDRNGTLIGOVDLKAKVNTLSTDIPEQLSKY
 VDNQRLSTFEYIKNIINTSILNIRESNLIDLSRASKINIGSKVNPIDKNOI
 QLFNLESSKIEVILKNAIVNSYENFSTWIRIPKFNISLNEEYTIINCMNNS
 GWKYSNLGELIWTLODQIEIKORVVEKYSOMISIDINMEIYITINNLSNKIY
 INGRILDKPISNGIHNASNNIEKLDGDFHRYIIXKFNLEKDLNKEEIDLY
 DNOSNGILDKPMGDYIOYKPYMLNLYDNKYVVDVNVNNGITRGMYIKGRSGMT
 NIYLNLSIRKTEKTIKKYASGNKNDIYRNDRYINVYKKEKRLTNASQAVEK
 ILSALEIPDVGNLSQVVVYKSKNDQGITNCKKMNIDNNGMDIGIFGHQNNIAKL
 ASNNYNRQIERSRSLTCSWEFIPVDDGMRPL"
 misc_feature
 /note="potential terminator: putative"
 4400..4432
 BASE COUNT 1934 a 517 c 756 g 1628 t
 ORIGIN

Query Match 54.8%; Score 727.2; DB 1; Length 4835;
 Best Local Similarity 72.2%; Pred. No. 1.6e-176;
 Matches 945; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

Db 3127 ATTGAGCTAATTTTAAAAAATGCTATTGTATATATAGTATGATGAATTTTACTACT 3186
 QY 259 tcccttcgtatccgtatcccgaaatctcaactcatctctctgtaacaataataacac 318
 Db 3187 ACCCTTTTGATGAAGATTCCTTAAGTATTTTAAACAGTATAGTCTAAATATGAATATACA 3246
 QY 319 atcaatcaactgtaagaacaacttcgttgtaagaatctctctgtaacgtgtaaac 378
 Db 3247 ATATATTAATTTGATGAAAAATATTTACAGATGGAAGCTATCACTTAATTTAGTGAATAA 3306
 QY 379 atctgtaactctgtaagacatcaagaatacaacagcgtgtgtatcaataactctcag 438
 Db 3307 ACTGTGACTTTTACAGATCTCAAGAAATATAAACAAGAGTAGTGTATTAATACACTCAA 3366
 QY 439 atgatacaactctctgtaacatcaatctgctgtatctctgtatccatccacaactgt 498
 Db 3367 ATGATTAATATATCAATATATATAACAATGATGATTTTGTATCTATTAATATACA 3426
 QY 499 ctgaataactccaaatctacatcaacgcgcgtctgatacgaacgaacccgatctcaat 558
 Db 3427 TTAATAACTCTAATAATTTATATTAATGGAAGATTAAATGATCAAAAACCAATTTCAAT 3486
 QY 559 ctgggtataacatcaacgcctcttaataacatcatgltcaaacctgtaacggtgtgtaac 618
 Db 3487 TTAGTAATATTTATCATGCTAGTAAATATATATATGTTAAATTAATAGATGTTAGATACA 3546
 QY 619 caccgtatcatctgtaacaaatctcaatctgcttgcgaagaactgtaacgaagaaga 678
 Db 3547 CATAGTATATTTGGATTAATAATTTTAACTTTTGTATGATGAAGAAATTAATGAAGA 3606
 QY 679 atcaagacctctgtaacacacacgcgtcaatctctgtaactctgtaaacactctggtgac 738
 Db 3607 ATCAAGATTTTATATGATATATCATCAATAATTCAGATTTTAAACACTTTTGGGTGAT 3666
 QY 739 taactgcaatgtaacaaacgcgtactacatgctgtaactctgtaacatccgaacaatacgt 798
 Db 3667 TTTTTCATATATGATTAACCATATATATATGTTTAAATTTATATGATCAATAATATATGTC 3726
 QY 799 gacgtcaacaatgtagtctccggttcaatgtaactgtaacgaagtcggtgtagtctgt 858
 Db 3727 GATGTAATATATGATATATGATATATGATATATGATATATGATATATGATATATGATAT 3786
 QY 859 atgataccaacatctcaactgtaactctccctgtaacggtgtaacgaactcatcaacag 918
 Db 3787 ATGACTACAAACATTTAATTAATTAATCAAGTTGTATAGGGGCAAAATTTATATATAA 3846
 QY 919 aaatacgcgtctgtaacaaagacaatctatctgtaacaaatgatactgtaatacat 978
 Db 3847 AATATATGCTTCTGGAATATGAATATATATGTTAGAAATATATGATGCTATATATTAAT 3906
 QY 979 gtgtgtagttaagaacaagaataacgctgctgtaacaaatgcttcaaggctgtgtagaa 1038
 Db 3907 GTATGATGTTAAATATGAATATATGATGTTAGCTTAAATGATGATGATGATGATGATG 3966
 QY 1039 aagatctctgctctgtaacaaatcccgagcgtgtgtaactctgtaacggtgtagttagt 1098
 Db 3967 AAATATCTAATGATGATTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4026
 QY 1099 aaatccaagaacgaacaggtatcaatacaaatgcaaaatgaaatctgtaagaacaacat 1158
 Db 4027 AAGTCAAAAAAATGATCAAGAAATTAATAATTAATTAATTAATTAATTAATTAATTAAT 4086
 QY 1159 ggtaacgataatggttcaactggttcccgagcttcaacaaatataatgtaactggtgt 1218
 Db 4087 GGAATATGATATATGCTTTATAGGATTTCAAGCTTTAATTAATTAATTAATTAATTAAT 4146
 QY 1219 tccaactgtacatcgttaagatcgtaacgttccctcgtaacgtggtgtagtctgtgtag 1278
 Db 4147 AGTAATTTGATATATATGCAAAATATGAAGAATCTAGTAGACTTTGGGTGCTCATGGGA 4206
 QY 1279 ttcataccggtgtgtagaagcgttggtgtagaagtcgcgctgtaagaatc 1326
 Db 4207 TTTATTTCTGTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4254

Db 2650 GAAATATCATTTAATAGACTTATCTAGTATGCATCAAAAATAATTTGTAGTAA 2709
 Qy 139 gtaacttcgactcgatcgacagaatcagatccgctggttcaactcgtgaactccca 198
 Db 2710 GTAAATTTTGTATGAAATGATTAATAAATCAATTAATTAATTTAAAGTGTAA 2769
 Qy 199 atcgaagttacccgaagaatgctatcgatatacaactatgtaagaactccacc 258
 Db 2770 ATTAGGATTAATTTTAAATAATGCTATGTATATATAGTATGTAATAATTTTGTACT 2829
 Qy 259 tcccttcgactcgatcccgaaatactccaactcctcgtgaacaatgatacc 318
 Db 2830 AGCTTTTGATGAAATTCCTTAAGATTTTAAACAGTATAGCTTAATTAATGAATATACA 2889
 Qy 319 atctcaactcgtggaagaacaactcgttggaagaatcctgtaactcgtgaact 378
 Db 2890 ATAAATTTTGTATGAAATTAATTCAGATGGAAGATATCATTAATTAATGTAATA 2949
 Qy 379 atctggaactcgtcgaagacactcgaagaaatcaaacagcgtgtgtatcaatactcag 438
 Db 2950 ATCTGACATTTACAGATACACGAAATTAACAAAGAGTGTGTTTAAATACAGTCA 3009
 Qy 439 atgatacaactcgtgactacatcaatcgtcgtgactcgttaccataccaactcgt 498
 Db 3010 ATGATTAATTAATCATGATTAATAAACAAGATGATTTTGTAACTATCATTAATAGA 3069
 Qy 499 ctgaaactcgaacatcaatcaacagcgcgtcgtacaccagaacccgactccat 558
 Db 3070 TTAATTAATCTGTAATAATTTATTAATAATTAATTAATTAATTAATTAATTAAT 3129
 Qy 559 ctgggttaacatccacgactcctaatacaatcgttcaactcgtgaactcgtcgtac 618
 Db 3130 TTAGGTATATTTACAGTATGATTAATAATTAATTAATTAATTAATTAATTAATTAAT 3189
 Qy 619 caccgtaactcgtgactacatcaactcgtcgtgaacgaactgaacgaaga 678
 Db 3190 CATGATTAATTTGATGATTAATTAATTTTGTAAAGATTAATTAATTAATTAATTAAT 3249
 Qy 679 atcgaagcctgtaacgaacagcgaactcgtgactcgtgaagaactcgtcgtgac 738
 Db 3250 ATCAAGATTTATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3309
 Qy 739 tactcgtcgtacgaacacgctactacatcgtcgaactcgtgaactcgaacacacg 798
 Db 3310 TATTTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3369
 Qy 799 gaactgaacatgtagtactcgtcgtgaactcgtgaagaactcgtcgtgactcgt 858
 Db 3370 GATGTAATTAATTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3429
 Qy 859 atgactgaacacatcactcgaactcctccgtcgtcgtgaactcgtgaacacacac 918
 Db 3430 ATGATTAATTAATTTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3489
 Qy 919 aaatacgcgtcgtgaacagaacatacgtcgaacacacacacacacacacacacacac 978
 Db 3490 AAATATGCTTCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3549
 Qy 979 gttgtagtgaagaac 1038
 Db 3550 GTAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3609
 Qy 1039 aagatctgctcgtcgtgaac 1098
 Db 3610 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3669
 Qy 1099 aaatccagaac 1158
 Db 3670 AAGTCAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3729
 Qy 1159 ggtgaac 1218

Db 3730 GGAATGATATAGCTTTATAGATTTCAATTCATTAAATATAGTAACACTAGACA 3789
 Qy 1219 tccaactgtaacatcgtgaac 1278
 Db 3790 ACTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3849
 Qy 1279 ttcaccccggttgatgaac 1320
 Db 3850 TTATTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3891

RESULT 12
 AX036248
 LOCUS AX036248 3891 bp DNA linear PAT 16-NOV-2000
 DEFINITION Sequence 27 from Patent EPI041149.
 ACCESSION AX036248
 VERSION AX036248.1 GI:11225866
 KEYWORDS
 SOURCE
 ORGANISM
 synthetic construct.
 REFERENCE
 1 (bases 1 to 3891)
 AUTHORS
 Kink, J.A., Ficca, J.R., Padhye, N.V., Thalley, B.S., Stafford, D.C. and
 Williams, J.A.
 TITLE
 Vaccine and antitoxin for treatment and prevention of C. Difficile
 disease
 JOURNAL
 Patent: EP 1041149-A 27 04-Oct-2000;
 PHARM PHARM INC (US)
 FEATURES
 source
 location/Qualifiers
 1..3891
 CDS
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 1..3891
 /note="synthetic"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC16480.1"
 /db_xref="GI:11225866"
 /translation="MOFVKNQFNRKDPVNGVDIATIKIPNGOMQPVAKFIHNKIW
 IPEIDTPEPREGDLPPEPKQVYVSYYDSYSTDENKNTYKGVKTLPEEYSTD
 IGRMLTSTVKGIPWGGSTIDTLEKAVIDTNCINVTDPDGSYRSEELVITIGSADI
 IOFEKSGFHEVLNLTTRNGYSTDYIIRSPDTEFEESLSEVFEELRFGGDAFIDS
 VLNHELTHAGRLYGLIATINPNRKYVTNAYEWSGLEVEFEELRFGGDAFIDS
 LOENEFRLYYNKEFDIATLNKASISYGTASLOMYNVEKREKILSEDSFGESVD
 KLRPKDKMLLEIYTEEDNPVPEFKYLRKTYLNDKAVKFINIYKVNATYDGFNL
 RNTNLAAMPNONTETINMNFETKLKNEFTEFEELKLCVRCITTSKTSLSKQVKNAL
 NDCLRVNMDLFFSPSEDNFTNDLNKGEELTSDNIAEENISLDLQOYLFENF
 DNEPENISIEMLSDIIGOLEMPNIEFPNGKYEIDKYTFHLRQDEHGSRI
 ALTNSVNEALNPSRVYTFPSSDYVKKNKATEAMPGLGVEOLYDPTDESEVST
 DKADITFIIRYIGPALNIGMLYKDEDFVGLIFSGAVILLEFIEIPIAIVGFAV
 SYLNKVLVTQTDINALSKRNEKDEVEYIVTNMLAVTNQIDILRKKMEALENOA
 EATKATINQYNTREEKNINININIDLSKINSIKKAMINIKPLNOSYSILAM
 SMIPGVRLDEDFDLSLDALIKYIDNRGLIGVDRLKRVNNTLSTDIPLFOLSKY
 VDNORLSTPEYIKNIINTISILNRYSNHLIDLSRASKINIGSKNFPIDKNOI
 OLFNLESEKIEVILKNAIVYNSMENEFTSWIRPKYFNSINLENYTIIINCENNS
 GKVSLNSELGIEMTODROEIKORVFEYSOMINISDVINISFVITINNNLSKIY
 INGRILDKPISLNGITHASNNIMKELGCDTHRYITKXFNPLEKELNKEIKDY
 DNGSNGILKPFMGEDYLDIYKPYTHLIDPNKYVDVNVNNGIRGMYIKRGSGVMTT
 NIVLNSSLYRGTKFLIKRYVSGKNDNIYRNDRYIINVVNNKERYLRTNNSQGVK
 ILSLAIEIDVGNLSQVVMKSKNDGINKKMNIDNNGNDIGFIQHPNNIATKLV
 ASNNYNRQIERSSRLTSGSWEFIVDDGMBGRPL"

BASE COUNT 1580 a 392 c 1290 t
 ORIGIN

Query Match 54.8%; Score 726; DB 6; Length 3891;
 Best Local Similarity 72.4%; Pred. No. 3.1e-176;
 Matches 942; Conservative 0; Mismatches 360; Indels 0; Gaps 0;
 Qy 19 ttactcttaccgtaacatcgaac 78
 Db 2590 TCTACATTTACGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2649

QY 79 gaatcaatcaatcgtatcgactcgtctcgtcagcttcacaaatcaacatcggtctctaa 138
 Db 2650 GAAAGTAATCAATTAATGACTTATGCTATGATGATCAATCAAAATTAATGTTAGTAA 2709
 QY 139 gtaactcgtatcgatcgacgaagaatcgatccagctgttcaatctgaactctccaa 198
 Db 2710 GTAATATTGATGATCAATTAATGATTAATTAATTAATTAATTAATTAATTAATTA 2769
 QY 199 atcgaagtatcctgaagaatcgtatcgatcaactcgtatcgaagaactctccac 258
 Db 2770 ATTGAGGTAATTTTAAATAATGCTATGTTATTAATGATGATGATGATGATGATGAT 2829
 QY 259 tccctcgtatcgatcgatccgaataactcaactcgtatcgtcgaagaatgaatacacc 318
 Db 2830 ACCTTTGGATTAAGTAATTCCTAGTATTTTAAACAGTATTAAGCTTAATTAATGATTA 2889
 QY 319 atcatcaactcgtatcgatgaagaatcgttggtaagaatcgtcgtcgaagaactcga 378
 Db 2890 ATTAATTAATTTGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2949
 QY 379 atctgactcgtcgaagacacatcgaagaatcgaagaactcgttgcgtatcgaatactc 438
 Db 2950 ACTGAGACTTTTACAGGATGATCTCAGGAATTAATAAACAAGAGTGTTTTAAATGACTCA 3009
 QY 439 atgatacaacatcgtcgtacacatcaatcgctgtgactcgttccatcccaacaatcgt 498
 Db 3010 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3069
 QY 499 ctgaataatcgaagaatcgaagaactcgaagaactcgtcgtcgaagaagaactcgaact 558
 Db 3070 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3129
 QY 559 ctgggtgacatccacacgtctcgaagaatcgaagaactcgtcgtcgtcgtcgtcgtcgt 618
 Db 3130 TTAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3189
 QY 619 caacgtatcgtatcgatcaactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 678
 Db 3190 CATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3249
 QY 679 atcaagaacatcgtcgaagaactcgaagaactcgtcgtcgtcgtcgtcgtcgtcgtcgt 738
 Db 3250 ATCAAAAGATTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3309
 QY 739 taactcgtatcgatcgaagaactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 798
 Db 3310 TATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3369
 QY 799 gacgtcaagaatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 858
 Db 3370 GATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3429
 QY 859 atgatacaacatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 918
 Db 3430 ATGACTTACAAAGATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3489
 QY 919 aaatacgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 978
 Db 3490 AATAATGCTTCTGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3549
 QY 979 gctgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1038
 Db 3550 GTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3609
 QY 1039 aagatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1098
 Db 3610 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3669
 QY 1099 aaatacgaagaactcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1158
 Db 3670 AAGTCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3729
 QY 1159 gatacgaatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1218

Db 3730 GGAATGATTAATGCTTTTATGATTTTCACTTTTAAATTAATTAATTAATTAATTAAT 3789
 QY 1219 tccatcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1278
 Db 3790 AGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3849
 QY 1279 ttcacccgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 1320
 Db 3850 TTTATTCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3891

RESULT 13

CBNTOXA

LOCUS CBNTOXA 4067 bp DNA linear BCT 24-JUN-1994
 DEFINITION C.bottulinum gene for infant neurotoxin type A.

ACCESSION X73423

VERSION X73423.1 GI:507070

KEYWORDS botulinum neurotoxin; botulinum neurotoxin type A.

SOURCE Clostridium botulinum.

ORGANISM Clostridium botulinum.

REFERENCE 1 (bases 1 to 4067)

AUTHORS Willems, A., East, A.K., Lawson, P.A. and Collins, M.D.

TITLE Sequence of the gene coding for the neurotoxin of Clostridium

TITLE botulinum type A associated with infant botulism: comparison with

TITLE other Clostridium neurotoxins

JOURNAL Res. Microbiol. 144 (7), 547-556 (1993)

MEDLINE 94143603

REFERENCE 2 (bases 1 to 4067)

AUTHORS Willems, M.C.A.

TITLE Direct Submission

JOURNAL Submitted (17-JUN-1993) A. Willems MCG, AFRC Institute of Food

JOURNAL Research, Dept of Microbiology, Reading Laboratory, Earley Gate,

JOURNAL Whiteknights Road, Reading, RG6 2EF, UK

FEATURES location/Qualifiers

source

RBS

CDS

misc_feature

1384..1386

Query Match 31.9% Score 423.6; DB 12; Length 1299;
 Best Local Similarity 62.0% Pred. No. 2.5e-98;
 Matches 817; Conservative 0; Mismatches 444; Indels 57; Gaps 7;

QY 6 cgaacagatgagccttaacttaactgaatatacaagaacatcaataactcaactcct 65
 DB 18 CAATAATCCGATACCTGATCTACACAACTGTAACAAAAATCAAGAACACTCTCTCT 77
 QY 66 gaacccgagctgaacgaatcaactgaatgaacccgctgctgaactgaactcaactca 125
 DB 78 GGACATGCGTTTACGAAACACAAATTCATCGACATCTGCGTATGTTTACATCTTC 137
 QY 126 catcggttctaaagttaacttcgatccgaatcgaaagaatcgaatcgaatcgaatc 185
 DB 138 TATCAACGGGTGACGCTACATCTACTACTAACCAGCAACGATGGTATGATCTCTTC 197
 QY 186 ggaatcttccaaatcgaagtatcctgaagaatgctatcgataaactctatgaaga 245
 DB 198 TAAACCGTGTGAAGTAAACATCGCTGAGAACAGACATCTCTACACGCTGTTACGA 257
 QY 246 aaactctcacctcctctctgataccgtaaccgaatcaactcaactcctcctga 305
 DB 258 GAACCTCTCTATCTCTTCTGCTGCTATCCGAATATCTTACAAAGTTAACTGAA 317
 QY 306 caatgaatacacatcaactcaactgcat---ggaacaaatctggttgaaagtatcct 362
 DB 318 CAACGAATACATCATGATGACTGATCGTAAACAACTCGTGGTGAATAATCTCTCT 377
 QY 363 gaactcagtgtaaatatctgactgaagacactcgaagaatcaagaagtggt 422
 DB 378 GAACCTCAACAAATATCTGATCTGACGACACTGCTGTGTAACAGAAACTGGT 437
 QY 423 attcaaatctcctgaatgaatcaacatcctgaactcaactcgtgactcgtctac 482
 DB 438 TTTCAACTACACTCAGATGATCTCTGATCTGATCTAATAAATGATCTGCTTAC 497
 QY 483 catcacaacaaatcgtcgaataactcaacaaatcaatcaacacgagcgtctgatac 542
 DB 498 TATCACTAACAAACCGCTGCTGATCTGATCTGATCAATCAACGGTGAACCTGAT 557
 QY 543 gaacccgactcgaatcgtgtaacatcgaactccttaataactatgattgaactga 602
 DB 558 AAATATATCTCTAACCTGGGTGATCCACAGTTCTGATGACAACTCTGTTCAAAATCG 617
 QY 603 cggctctcgaacactcgaactcgaactcgaactcaactcaactcgtctgatacga 662
 DB 618 TGGTTCCAAACGACAC---GCGTTAGCTTGGTATCCGTTACTTCAAGTTTGCACACTGA 674
 QY 663 actgaacgaacaaagaatcaagaacctgtaagaacacgagtcgaatctgatacctga 722
 DB 675 ACTGGGTAAACCTGAATGAAACTCTGATCTGATGACGAACCGACCGCTATCTCTGAA 734
 QY 723 aaactctgagtgatgaactcgaatgaagaacacgctgaactcgaatcgaatcgaatga 782
 DB 735 AGACTCTGGGTAACTACTGCTGTAACAACAGTTACTACTCTGAAACCTGCTCCG 794
 QY 783 tccgaacaaatacgtgagctcaacaatgagtlacccggtgataactgaactgaaga 842
 DB 795 GACTGACAAATCTATACACAGACTCT-----AACTTCTTGAACATCAACCA 842
 QY 843 tccggtggtctgatagaactcaacatcgaactcctcctgatacgggtgac 902
 DB 843 GGAGCGTGTGTTTATCAAAACCTTAATCTCTTAACACACTGCTGTACACTGCTG 902
 QY 903 caaatcgaatcaagaatgaagctct-----ggttaacaagaacaaatcgtctg 953
 DB 903 TGAAGTTATCATCCGTAATAACGTTCTACTGACATCTCTACACACTGCAAACTCTGAC 962
 QY 954 caacaatgatacgtgatacaatcaatgctgagttgaagaacaaagaatcagctgctac 1013
 DB 963 TAAAAACGACCTGGCTTACATCAAGCTTGTGACCGTGACGTTGAATACCGTCTGTACCG 1022

QY 1014 caatgcttcagagctggtgataagaatctgtctcgtcgtgaatcccgagctg 1073
 DB 1023 TGACATCTCTATGCTTAACCGGAAATAATCATCAACATGATCCGTAATCTTA 1082
 QY 1074 taa---tctgctcaggtgagtgataagaatccaagaacgacgaggtataactaca 1130
 DB 1083 CAACCTCTGGGTCAGATCATCTGTTATGACATC-----GATCGGTACAA 1127
 QY 1131 atgcaaatgaatctgcaagacacaaatgtaacgaatcgttctcgttccacaa 1190
 DB 1128 CTGCACTATGAACTTCCAGAACACAGCGTGTATACATCGTCTGCTGGTTCCACTTC 1187
 QY 1191 gtccaacaalatgcgttaaacctggtctcccaactggtacaaatcgtcgaatcgaactc 1250
 DB 1188 TAAACAAC-----CTGCTGCTCTCTTCTTGTGTAACACACATCCGTAAAA 1235
 QY 1251 cttctcagctcgtggtgctcgttggaagttcaccggtgataagaggttgagga 1308
 DB 1236 CACTTCTTAAACGGTGTGCTGCTGCTTCTTCACTCTTAAAGAACACGCTTGGCAGGA 1293

RESULT 15
 LOCUS A58946 1313 bp DNA linear PAT 06-MAR-1998
 DEFINITION Sequence 6 from Patent WO9641881.
 ACCESSION A58946
 VERSION A58946.1 GI:3714418
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 1313)
 AUTHORS Elmore,M.J., Mauchlille,M.L., Minton, Nigel,P. and Pasechnik,V.A.
 TITLE TYPE F BOTULINUM TOXIN AND USE THEREOF
 JOURNAL Patent: WO 9641881-A 6 27-DEC-1996;
 MICROBIOLOGICAL RES AUTHORITY (GB)
 COMMENT Other publication AU 6012196 970109.
 FEATURES
 source 1..1313
 /organism="unidentified"
 /db_xref="taxon:32644"
 BASE COUNT 401 a 344 c 210 g 358 t
 ORIGIN

Query Match 31.8% Score 422; DB 6; Length 1313;
 Best Local Similarity 61.9% Pred. No. 6.6e-98;
 Matches 816; Conservative 0; Mismatches 445; Indels 57; Gaps 7;

QY 6 cgaacagatgagccttaacttaactgaatatacaagaacatcaataactcaactcct 65
 DB 26 CAATAATCCGATACCTGATCTACACAACTGTAACAAAAATCAAGAACACTCTCTCT 85
 QY 66 gaacccgagctgaacgaatcaactgaatgaacccgctgctgaactgaactcaactca 125
 DB 86 GGACATGCGTTTACGAAACACAAATTCATCGACATCTGCGTATGTTTACATCTTC 145
 QY 126 catcggttctaaagttaacttcgatccgaatcgaaagaatcgaatcgaatcgaatc 185
 DB 146 TATCAACGGGTGACGCTACATCTACTACTAACCAGCAACGATGGTATGATCTCTTC 205
 QY 186 ggaatcttccaaatcgaagtatcctgaagaatgctatcgataaactctatgaaga 245
 DB 206 TAAACCGTGTGAAGTAAACATCGCTCAGAACAGACATCATCTACACGCTGTTACGA 265
 QY 246 aaactctcacctcctctcgtgataccgtaaccgaatcaactcaactcctcctga 305
 DB 266 GAACCTCTCTATCTCTTCTGCTGCTGATATCCGAATATCTTAAACAACTTAACTGAA 325
 QY 306 caatgaatacacatcaactcaactgcat---ggaacaaatctggttgaaagtatcct 362
 DB 326 CAACGAATACATCATGATGACTGATCGTATCAACAACTCGTGGTGAATAATCTCTCT 385

Qy 363 gaactacggttgaaatcaltctggaactctgcaggacactcaggaatcaacaacgctgtgt 422
Db 386 GAACCTACACAAATCATCTGAGCTCTGAGGACACTGCTGTACACACAGAACTGCT 445
Qy 423 attcaaatctctcagaatgatacaactctgtaactacaatcgcctgtgactctgttac 482
Db 446 TTTCAACTACACTCAGATGATCTCTACTACTACATTAATAATGATGATCTGCTTAC 505
Qy 483 catcaccaaatctcgtgaatcaactccaatctacatcaacgagcgctctgatcgacca 542
Db 506 TATCACTAACACCGCTCTGGGTACTCTGATCTACATCAACGGTAACTGATCATGA 565
Qy 543 gaaacggaatcccaatctggtgtaacatccacgctctctaaataacatgltcaaatgta 602
Db 566 AAAATCTATCTTACTACCTGGGTGATCCAGCTTCTGACAACTCTTCAAAATGCT 625
Qy 603 cgggtgtcgtgaacactcaccgctacatctgatacaaatcaatctgttcgacaaga 662
Db 626 TGGTTGCAACGACAC---GCGTTACGTTGGTATCCGTTACTTCAAAAGTTTTCGACACTGA 682
Qy 663 actgaacgaaataaataaataaagactctgagacaacacgctcaatctcgtatctgaa 722
Db 683 ACTGGGTAAACTGAAATCGAATCGAATCTGACTCTGACGACCGACCGCTTATCTCTGAA 742
Qy 723 agactctcgtggtgtaactactcgaagtaacgacaacgctactacatgctgaatctgacga 782
Db 743 AGACTTCTGGGGTACTACTCTGCTGACACAAACGTTACTACTGCTGACCTGCTCG 802
Qy 783 tccgaacaaatacgttgacgtcaacaatgtaagtaacggtttacatgttaccctgaag 842
Db 803 GACTGACAAATCTATCACTCAACACTCT-----AACTTCTGAACTCAACCA 850
Qy 843 tccggtgtgtctgtatgtaactcaacatctcactccttccctgtacggtgtac 902
Db 851 GCAGCGTGTGTTTATCGAATACTTAATATCTTCTTCTATCACTCGTCTGTACACTGCTGT 910
Qy 903 caaatlcatlcaagaataacgctct-----ggtacaagaacataatcgtctg 953
Db 911 TGAAGTATCATCCGTAATAAAGGTTTCTACTACATCTCTAACACTGACACTGACACTGTG 970
Qy 954 caacaatgatacgtatatacatcaatgtgtgtaagtaagaacaaataacgctctggtac 1013
Db 971 TAAAAACGACCTGCTTACATCAACGTTGTGACCGTGAATGATACCGTCTGTACGC 1030
Qy 1014 caatgctctcaggtgtgtgtaagaatctgtctgctcgtgaatccggaagtgtg 1073
Db 1031 TGACATCTCTATCGCTAAACCGGAAAAATCATCAACTGATCCGTACTTCTAACTCTAA 1090
Qy 1074 taa---tctgtctcaggtgtgtgtaatgaatccaagaacgacggtatcatcaaca 1130
Db 1091 CAACTCTCTGGGTGATCATCTGATGACTC-----GATCGGTACAA 1135
Qy 1131 atgcaaatgaaatctgcaggaacaaatgtaacgatatcggtttcaatcggtttccaca 1190
Db 1136 CTGCACTATGAACCTTCAGAACACACACGCTGTGTAACATCGCTGCTGTTCCACTTC 1195
Qy 1191 gtccaacatatcgtaaatcgtgtctcgaactcgaatcgtaacatcgtaacgaatc 1250
Db 1196 TAACAAAC-----CTGTTGCTTCTTCTTGTGTAACAACTGATCCGTAATAA 1243
Qy 1251 ctctcgcaactcgtgtgtctctctggaagtltacccggttgaatgacggttgggttga 1308
Db 1244 CACTTCTTCTTAACGCTGCTCTGCTTCTTCAATCTTAAGAACAACGCTTGGCAGGAA 1301

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 19:28:11 : Search time 4365.92 seconds
(without alignments)
4089.968 Million cell updates/sec

Title: US-09-611-419a-3

Perfect score: 1323
Sequence: 1 gaatcgaacagatgctac.....aacgtccgtctaagaattc 1323

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.4	4.4	608	10	BI887904 2F637-1-0
2	54.6	4.1	1101	12	AL068607 Drosophila
3	53.6	4.1	537	10	BM277996 AS_tgz-51
4	53.6	4.1	551	10	BM278689 AS_tgz-66
5	53.2	4.0	394	9	AU060224
6	52.6	4.0	447	10	BI594803 AS_tgz-44
7	52	3.9	597	10	BM278181 AS_tgz-54
8	51.8	3.9	500	10	BM278805 AS_tgz-67
9	50.6	3.8	511	10	BM004465 TGESTZya8
10	50.4	3.8	569	10	BM278657 AS_tgz-65
11	50	3.8	641	10	BM181884 fvs1b1.y
12	49.6	3.7	645	9	AI389106 GH20192.5
13	49.2	3.7	827	12	CNS02156 Tetradon
14	49	3.7	481	10	BF050030 AS_tgz-21
15	48.8	3.7	628	9	AU060230
16	48.2	3.6	546	10	C22974 C22974 Dict
17	48.2	3.6	1101	12	CNS0100X

18	48	3.6	546	10	BF169335
19	48	3.6	556	10	BM277853 AS_tgz-49
20	47.4	3.6	377	10	BI594814 AS_tgz-44
21	47.4	3.6	450	9	AU060996
22	47	3.6	458	10	BI783045
23	47	3.6	518	10	BI782765
24	46.8	3.5	1175	10	BI872945
25	46.4	3.5	469	10	BF050033
26	45.4	3.4	386	10	BF050064
27	45.4	3.4	439	10	BM278692
28	45.4	3.4	536	10	BF050073
29	45.4	3.4	542	10	BM278786
30	45.2	3.4	554	10	C23753
31	45.2	3.4	501	10	BM277916
32	45.2	3.4	525	12	BM184460
33	45.2	3.4	525	12	CNS07PEA
34	45.2	3.4	1101	12	CNS000D1
35	45	3.4	564	10	BE060733
36	44.8	3.4	534	10	BM278087
37	44.6	3.4	1101	12	CNS016JY
38	44.4	3.4	522	10	BM278558
39	44.4	3.4	681	12	CNS02EOD
40	44.2	3.3	325	12	CNS03PK8
41	44.2	3.3	429	9	AU052930
42	44.2	3.3	500	12	B67199
43	44.2	3.3	519	10	BM278164
44	43.8	3.3	535	10	BI594948
45	43.6	3.3	443	10	BM039802

ALIGNMENTS

RESULT 1

BI887904 608 bp mRNA EST 12-OCT-2001
LOCUS 2F637-1-002159 zebrafish shield stage whole embryo cDNA library
DEFINITION MPMGP637 Danto rerio cDNA clone MPMGP637_10F4:MPMGP637F0410 5',
mRNA sequence.

ACCESSION BI887904.1 GI:16095175
VERSION
KEYWORDS
SOURCE
ORGANISM Danto rerio

REFERENCE
AUTHORS Clark,M., Anstad,P., Hennig,S., Johnson,S.L. and Lehrach,H.
TITLE 1 (bases 1 to 608)
JOURNAL Unpublished (2001)

COMMENT
EST sequencing of a zebrafish shield stage cDNA library normalised by oligonucleotide fingerprinting
Unpublished (2001)
Contact: Hennig S
Laboratory 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inhestr.63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1360
Email: hennig@molgen.mpg.de
5' EST sequencing of clones from a zebrafish shield stage library, normalised from 55,000 starting clones by oligonucleotide fingerprinting
High quality sequence stop: 608.

FEATURES
source location/Qualifiers
1. 608

/organism="Danto rerio"
/db_xref="taxon:7955"
/clone="MPMGP637_10F4:MPMGP637F0410"
/clone.lib="zebrafish shield stage whole embryo cDNA library MPMGP637"
/tissue_type="whole embryo"
/dev_stage="shield stage, 6 hrs post-fertilisation"
/lab_host="E.coli, X11 blue MRF"

BASE COUNT	ORIGIN
207 a	218 c
	41 g
	140 t
	2 others

Query Match	4.48;	Score 58.4;	DB 10;	Length 608;
Best Local Similarity	49.18;	Pred. No. 0.00033;		
Matches 181; Conservative	0;	Mismatches 187;	Indels 1;	Gaps 1.

QY	439	atcaacatcctcgagctatacctaa-ccgcctggagctcttcgttaccatccaaacaaatcgctct	497
Db	5	ATCAACTTCTCGAGTACTACATCACTATCTCCAGCTACATCAACTTCTCCAAACAACAACA	64
QY	498	gaataactccaanaatctacatcaaaagcgccgtctgtatcgcacgaacacgattccaatct	557
Db	65	CACAAACTTCTCGAGCTTCATCAACTTCTCCAAACAACAACAACAACAACAACAACA	124
QY	558	gggtacatccacgccttctaataacatcaatgtctcaaacctgacggctgttcgtgacatca	617
Db	125	CATCAACTTCTCCAGCTTCATCAACTTCTCCAGCTACATCAACTTCTCCAGCTACAAACA	184
QY	618	ccgcctacatcttgatccaaataacttaatatctgttcgacaagaagacgaaagaaanaaat	677
Db	185	CTTCTCCAGCTACATCAACTTCTCCAAACAACAACAACAACAACAACAACAACAACA	244
QY	678	caaaagacctgtatcgacaacaccagttccaatctgtgtatctccgaaagaccttcggggtacta	737
Db	245	CACAAACTTCTCGAGCTTCATCAACTTCTCCAAACAACAACAACAACAACAACAACA	304
QY	738	ccctgcagtagcgaacaaaccgctactacatgcctgaatctgttaacgtccgaacaaatacgttga	797
Db	305	CATCAACTTCTCCAGCTTCATCAACTTCTCCAGCTACATCAACAACAACAACAACAACA	364
QY	798	cgtaacaaa 806	
Db	365	CTACACAACA 373	

RESULT	2
CNS00100	
LOCUS	1101 bp DNA linear GSS 03-JUN-1999
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC: BACR32D23 of RpCl-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL068607
KEYWORDS	AL068607.1 GI:4958689
SOURCE	GSS.
ORGANISM	fruit fly. Drosophila melanogaster Euarctota, Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101) Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr) - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutyo Osoegawa and Aaron Mamosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RpCl-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw ^{sp} , the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	

filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

location/Qualifiers

1. 1101

orientation: "Bacpacphila_malvector"

BASE COUNT	155 a	166 c	7 g	284 t	489 others
ORIGIN					

Query Match	4.1%	Score 54.6	DB 12,	Length 1101,
Best Local Similarity	17.6%	Pred. No. 0.0039		
Matches 100,	Conservative 249,	Mismatches 245,	Indels 3,	Gaps 1

[illegible]

RESULT	3
BM277996	
LOCUS	
DEFINITION	BM277996 537 bp mRNA linear EST 20-DEC-2001
ACCESSION	As_tgz_51F02-SKPL Ascaris suum adult male testis germinal zone from
VERSION	Aln_Scgt Ascaris suum cDNA clone As_tgz_51F02 5', mRNA sequence.
KEYWORDS	BM277996.1 GI:17971254
SOURCE	EST.
ORGANISM	pig roundworm.
	Ascaris suum
	Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarida; Ascaridoidea
	; Ascarididae; Ascaris.
REFERENCE	1 (bases 1 to 537)
AUTHORS	Blaxter, M.L., Parkinson, J., Whitton, C., Daub, J., Gulliano, D., Hall
	, N., Quayle, M. and Barrell, B.

TITLE Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL Unpublished (2000)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JF, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk

The library was prepared by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore. Sequencing was performed by the Pathogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell).

PCR primers

FORWARD: T3
BACKWARD: T7PL
Plate: 51 row: F column: 02
Seq primer: SKPL
High quality sequence stop: 478.

FEATURES
source location/Qualifiers

1..537
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As_tgz_51F02"
/clone_lib="Ascaris suum adult male testis germinal zone from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI; Library was made from dissected testis germinal zone from adult male Ascaris suum collected from abattoirs. Constructed by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore, MD."

BASE COUNT 193 a 149 c 72 g 123 t
ORIGIN

Query Match 4.1%; Score 53.6; DB 10; Length 537;
Best Local Similarity 48.6%; Pred. No. 0.0053;
Matches 180; Conservative 0; Mismatches 184; Indels 6; Gaps 1;

OY 359 tgaactcagtggaatcatctggaactctgcaggacatcgaagaatcaacacgctgtg 418
DB 49 TGTACACCAATGTCTCAACAGAAACCCACTGACAACTTCACACAGCAACTTTTA 108
OY 419 tatcaaatctctagatgatcaacatctctgactaatcaatcgctggtatctcgta 478
DB 109 GCAACTTCACAGTAACCTCAACACAGCCATTTCACACAGCAACCTCAACTCAATA 168
OY 479 ccataccacaacatcgctgtaatactcaaatctacatacgaagcgctgcatcgacc 538
DB 169 ACAACCTTAGCAACCTCTGACAACTTCATTTTTCGCAACTCTCTGAGACACG 228
OY 539 agaacgcgatctcaatctggttaacatccacgctcttaataatacatcgttcaaatg 598
DB 229 GGAG-----CTACCAACTTGAACAACCTCAACACCTCAATATCAACCTGATCTGA 282
OY 599 acggtgtcgagacatcagcgtacatctgataataataacttaacttggtcgacaag 658
DB 283 ACAACAATCTCAGCAATCTCAACAATTAACCTTCACAACTTGAACAGCAACTTCACA 342
OY 659 aactgaagaaagaatacaagaaccgtgacgaacacgctcaatctcgtgtatcctga 718
DB 343 ACAGTAATCTCAGCAACTTCACACAGCAACCACTTGAACACTTCAACAGCACTTA 402
OY 719 aagactctg 728
DB 403 ACAACAGCAG 412

RESULT 4
BM278689

LOCUS BM278689 551 bp mRNA linear EST 20-DEC-2001
DEFINITION As_tgz_66A05_SKPL Ascaris suum adult male testis germinal zone from Alan Scott
ACCESSION BM278689
VERSION BM278689.1 GI:17971947
KEYWORDS EST.
SOURCE pig roundworm.
ORGANISM Ascaris suum
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascarididae; Ascarididae; Ascaris.

REFERENCE 1 (bases 1 to 551)
Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guiliano,D., Hall,N., Quayle,M. and Barrell,B.
Edinburgh University/Sanger Centre Nematode EST Project
Unpublished (2000)

TITLE Contact: Blaxter ML
JOURNAL Institute of Cell, Animal and Population Biology
COMMENT University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JF, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk

The library was prepared by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore. Sequencing was performed by the Pathogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell).

PCR primers
FORWARD: T3
BACKWARD: T7PL
Plate: 66 row: A column: 05
Seq primer: SKPL
High quality sequence stop: 510.

FEATURES
source location/Qualifiers

1..551
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As_tgz_66A05"
/clone_lib="Ascaris suum adult male testis germinal zone from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI; Library was made from dissected testis germinal zone from adult male Ascaris suum collected from abattoirs. Constructed by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore, MD."

BASE COUNT 208 a 160 c 64 g 119 t
ORIGIN

Query Match 4.1%; Score 53.6; DB 10; Length 551;
Best Local Similarity 49.6%; Pred. No. 0.0053;
Matches 137; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

OY 428 actccaagatgtaacatctctgactaatcgcgtgatactctggtaccatcacca 487
DB 102 ACAGTAATCTCAACAACCAATTTTCACAGCAACCTCGCAACTTCATATCAACCTTA 161
OY 488 acaatcgtctgataatactcaaatctacatacaacgcccgtctgatacgacgaaccca 547
DB 162 GCAACTCTCTGAAACAATCTCAATAACAACCTCGCAACTCTGTGAGCAACCTCACTACA 221
OY 548 tctcaatctggtgtacatccacgctcttaataataataatcgttcaaatcgtgagctgtc 607
DB 222 ACTTGACCAACCTCAACAACCTTCACACCAACCTCAACTATCTCAATTAACATTCACGA 281
OY 608 gtgacacacacgctacatctgataataataactcaatcgttgcacaagaagtaagc 667
DB 282 GTCTCAACAATTAACCTCAACAACCTTAGCAACAGCAAGCAATTTTCAACAGCACTTA 341
OY 668 aaaaagaatcaagaacctgtgacgaacacagctta 703

DB 342 ACAGTACTTCAGCACTTCACAGCAACACTTCA 377

RESULT 5
LOCUS AU060224
DEFINITION AU060224 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
ACCESSION AU060224
VERSION AU060224.1 GI:4881328
KEYWORDS EST
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
AUTHORS 1 (bases 1 to 394)
Moriya, R., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
Yoshino, R., Mitra, B.N., Pi, M., Saito, T., Takemoto, K., Yasukawa, H.,
Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
TITLE Developmental cDNA in Dictyostelium discoideum
JOURNAL Unpublished (1998)
COMMENT Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Tennodai, Tsukuba, Ibaraki 305, Japan
Email: d402huesakura.cc.tsukuba.ac.jp
PROJECT - Dictyostelium discoideum cDNA project in Japan.

FEATURES
source
1. 394
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLA610"
/clone_11b="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"
BASE COUNT 223 a 134 c 8 g 28 t 1 others
ORIGIN

Query Match 4.0%; Score 53.2; DB 9; Length 394;
Best Local Similarity 49.3%; Pred. No. 0.0059;
Matches 139; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 423 caaatctctcagatgatacaactctctgatacatcaatcgctgattcgtaccat 482
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 111 CAACAACACCAACACACCAACCAACCAACCAACCAACCAACCAACCAACAA 170
QY 483 caccacaactcgtctgataaactcaaatctacatcaacggcgtctgatacgcagaa 542
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 171 CAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACAA 230
QY 543 accgatctccaatctggtatacaaccagctcttaataacatcatgttcaactggagc 602
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 231 TATCATCAACATCAGCAACAAATATCATATATCAACAAACAAATATCATCAACATCAACAA 290
QY 603 ttgtctgacacaccgcgtctacatctgatacattcaactcgtttcgacaagaact 662
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 291 CAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACAA 350
QY 663 gaacgaanaagaatcaagaacctgacgaacacagttcaa 704
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 351 CAACAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACAA 392

RESULT 6
LOCUS B1594803 447 bp mRNA linear EST 07-SEP-2001
DEFINITION AS_tgz_44C11.SKPL Ascaris suum adult male testis germinal zone from
ACCESSION B1594803
VERSION B1594803.1 GI:15498290
KEYWORDS EST.
SOURCE pig roundworm.
ORGANISM Ascaris suum

REFERENCE 1 (bases 1 to 447)
AUTHORS Blaxter, M.L., Parkinson, J., Whittton, C., Daub, J., Gulliano, D., Hall,
N., Quayle, M. and Barrell, B.
TITLE Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL Unpublished (2000)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared by Michelle Lizotte-Waniewski for Alan
Scott, Johns Hopkins University Medical School, Baltimore.
Sequencing was performed by the Pathogen Sequencing Unit, Sanger
Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
sequence contained a Polya tail (trimmed)
PCR Primers
FORWARD: T3
BACKWARD: T7PL
Plate: 44 Row: C Column: 11
Seq primer: SKPL
High quality sequence stop: 447.
Location/Qualifiers

FEATURES
source
1. 447
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As_tgz_44C11"
/clone_11b="Ascaris suum adult male testis germinal zone
from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/note="Vector: Lambda Zap II. Site 1: EcoRI, Site 2: XhoI;
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Lizotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."

BASE COUNT 173 a 127 c 49 g 98 t
ORIGIN

Query Match 4.0%; Score 52.6; DB 10; Length 447;
Best Local Similarity 49.8%; Pred. No. 0.0088;
Matches 133; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

QY 437 tgatacaatctctgatacatcaatcgctgattcgttaccatcacaacatcgtc 496
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 3 TCACAACACAGCAATTTTCACAGCAACCTCAGCAACTTCAATACCAACTTACCAACTCTC 62
QY 497 tgaatactccaaatctacatcaacggcgtctgatacgcagaacacgatctcaac 556
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 63 TGAACAACCTTCAATACCAACCTTACCAACTCTTGAGCAACCTCAGCTACCACTTGAACA 122
QY 557 tgggtacatccacgctcttaataacatcatgttcaactggagcgttgcgtgacac 616
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 123 ACCTCAACAACCTCAACCAACCACTCAATATCAATCAATCAATCAATCAATCAATCAATCA 182
QY 617 accgctacatctgatacaatctcaatcgttgcagaagaacacgacgaagaaga 676
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 183 ATAACTCAACCACTTACCAACCACTTCAACAGCAAGTAACCTTCAACCAACCAAGTAAC 242
QY 677 tcaagacacctgacgaacacagttca 703
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
DB 243 TCAGCAACTTCAACACCAACCACTTCA 269

RESULT 7
LOCUS BM278181 597 bp mRNA linear EST 20-DEC-2001
DEFINITION AS_tgz_54C05.SKPL Ascaris suum adult male testis germinal zone from

ACCESSION	Atlan Scott Ascaris suum cDNA clone As_tgz_54C05 5', mRNA sequence.
VERSION	BM278181
KEYWORDS	BM278181.1 GI:17971439
SOURCE	EST.
ORGANISM	pig roundworm.
REFERENCE	Ascaris suum
AUTHORS	Eukaryotes; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea 1 (bases 1 to 597) 1 ; Ascarididae; Ascaris.
TITLE	N., Quaghe, M., Parkinson, J., Whitton, C., Daub, J., Gulliano, D., Hall
JOURNAL	Edinburgh University/sanger Centre Nematode EST Project
COMMENT	Unpublished (2000) Contact: Blaxter ML

Asthorth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared by Michelle Lizotte-Maniowski for Alan Scott, Johns Hopkins University Medical School, Baltimore.
Sequencing was performed by the Patrogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell).
PCR Primers
FORWARD: T3
BACKWARD: T7PL
Plate: 54 row: C column: 05
Seq primer: SKPL
High quality sequence stop: 502.

```

/organism="Ascaris suum"
/db.xref="taxon:6253"
/clone="AS_tgz_54C05"
/clone_id="Ascaris suum adult male testis germinal zone
from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/note="Vector: Lambda zap II; Site 1: EcoRI; Site 2: XhoI;
library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abbaecois.
Constructed by Michelle Liscotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."

```

Query Match	3.9%	Score 52	DB 10	Length 557
Best Local Similarity	49.38%	Pred. No.	0.014	
Matches 136	Conservative	0	Mismatches 140	Indels 0
			Gaps	0

Qy	488	acaaatcgctcgtgaataaactccaanaatctacataccaagcgctctgtagtcagcaagaaccca	547
Db	201	gcaactctcttcaaacacacttcaatpaaacacccctcagcaacactgctgagcaaacctcagtttca	260
Qy	548	tctccaatctcgtgtaaacctcagcgtctctaaatacatgatgttcaactgtgaaggtgtc	607
Db	261	acttgamaacacctcacaacacctcattaaacacacttaattctcaacaacattctcaca	320
Qy	608	gtgaacactcaacgctacatctcgtgaatcaaatcttaacttcgtcttcgacaagaagaactgaa	667
Db	321	atfctcaacaaatpaaactcacaacacttaccagacagaaacttcaacagcagtaacttcamca	380
Qy	668	aaaaagaatatcaagaactcgttagcacaaccgctcca	703
Db	381	acagtaacttcagcaacttcaacagcaacaacacttca	416

RESULT	8
BM278805	
LOCUS	500 bp mRNA linear EST 20-DEC-2001
DEFINITION	As_tg-67F10.SKPL Ascaris suum adult male testis germinal zone from
ACCESSION	AJ269257
VERSION	BM278805
KEYWORDS	BM278805.1 GI:17972063
SOURCE	EST.
ORGANISM	Pig roundworm. Ascaris suum Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea ; Ascarididae; Ascaris. 1 (bases 1 to 500)
REFERENCE	Blaxter M.L., Parkinson J., Whitton C., Daub J., Gulliano D., Hall
AUTHORS	N., Quayle M. and Barrell B
TITLE	Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL	Unpublished (2000)
COMMENT	Contact: Blaxter ML

Armstrong Labs, King's Buildings, West Mains Road, Edinburgh, EH9
 3JF, UK.
 Tel: +44 131 650 6760
 Fax: +44 131 670 5450
 Email: mark.blaxter@ed.ac.uk
 The library was prepared by Michelle Lizotte-Manlewski for Alan
 Scott, Johns Hopkins University Medical School, Baltimore.
 Sequencing was performed by the Pathogen Sequencing Unit, Sanger
 Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
 sequence contained a Polya tail (trimmed)
 PCR primers
 FORWARD: T3
 BACKWARD: T7PL
 Plate: 67 row: F column: 10
 Seq primer: SKPL
 High quality sequence stop: 500.

```

/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="AS_tg-67F10"
/clone_1lb="Ascaris suum adult male testis germinal zone
from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI;
library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Litotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."

```

Query Match	3.9%	Score 51.8	DB 10	Length 500
Best Local Similarity	50.8%	Pred. No. 0.015		
Matches 153	Conservative	0	Mismatches 142	Indels 6
				Gaps 1

Accession	Sequence	Position
Oy	actctcaagatgatacaacatctctgacatcacaatcgttgatctctgttaccatcaaca	487
Oy	428 actctcaagatgatacaacatctctgacatcacaatcgttgatctctgttaccatcaaca	487
Db	70 ACAGTAACCTTCACACACAGCAATTTCAACAGCAACCTCAGCACTTCAATTAACAACCTTA	129
Oy	488 acaatcgtctgaataaactccaagaatctacatcaacagcgctctgataccagaaccca	547
Db	130 GCAACTCTCTGTAAMCAACTTCAATATAAACCTCCAGCAACTCTCTGAGCAAACTTCAG----	184
Oy	548 tctccaatctggtgaacatccacgctcttaataatcatcatgttcaaatgtagcggtgtc	607
Db	185 -CTAAACTTTGAACAACTCTCAACAACTTCATTAACAACTTATTCAAACAACATC	243
Oy	608 gtgacactacacggtacactctgatacaaatctcaatctgtcttcgacaaagaactgagc	667

[illegible]

RESULT	9
BM004465	
LOCUS	BM004465 511 bp mRNA linear EST 25-OCT-2001
DEFINITION	TgESTzya86f03.y1 TgVEG Partially sprulated oocyst cDNA Toxoplasma
ACCESSION	gondii cDNA clone TgESTzya86f03.y1 5' mRNA sequence.
VERSION	BM004465
KEYWORDS	BM004465.1 GI:16439239
SOURCE	EST.
ORGANISM	Toxoplasma gondii.
	Toxoplasma gondii

REFERENCE
1 (pages 1 to 511)
AUTHORS
Tang, K., Cole, R., Fogarty, S., Shipley, L.D., Ajioke, J.A., White, M.,

TITLE	Toxoplasma EST Project
JOURNAL	Unpublished (2001)
COMMENT	Contact: Clifton, S.

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: toxowatson.wustl.edu
Contact David Sibley (toxowest@orcin.wustl.edu) for further
information relating to organism, libraries, or clone availability
High quality sequence stop: 433.

FEATURES	Location/Qualifiers
source	1. 511

```

/organism="Toxoplasma gondii"
/strain="VEG (Type III)"
/db_xref="taxon:5811"
/clone="TGESRzYa86f03.y1"
/clone_lib="tgVEG Partially sporulated oocyst cdna"
/dev_stage="Partially sporulated oocysts"

```

/note=Vector: Modified pBluescript (pBSSK⁺) Site.1: BamHI Site.2: EcoRI. PolyA mRNA from partially-sporulated oocysts was converted to cDNA using the template-switching PCR method (SMART cDNA, Clontech Inc.,) and sized selected on sizeprep 400 columns (Amersham Pharmacia Biotech Inc.). First strand was reverse transcribed using the CDS III-oligo-dT primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified pBluescript vector containing directional SfiI sites, and electroporated into DH10B or DH2S cells. Vector: SfiI sites were added to the multiple cloning region of pBluescript SK⁺ between the BamHI/EcoRI sites. The modified polylinker has the following sequence:

Source: Michael White, Maria E. Jerome, Emily A. Johnson, Jay A. Radke, Montana State University. Clone Availability: David Sibley, Washington University"

BASE COUNT	213 a	181 c	61 g	56 t
ORIGIN				

Query Match	3.8%	Score 50.6	DB 10	Length 511
Best Local Similarity	46.9%	Pred. No. 0	03	
Matches 158, Conservative	0	Mismatches 179	Indels 0	Gaps 0

Oy	423	caaatnactctcgaatgatctgaacaactctctcgaacaaactcgaatctcgttaacct	482
Db	169	CTACTACTTCAACAAGTACGACACCAACAAACAACACTTCAACAAGCACTCAACACCAACAA	228
Oy	483	cacccaacaactcgtctcgaataactccaataatctacatcaacggcgcgtctcgtatcgaccagaa	542
Db	229	CAACGACCAACAACCAACCAACCACTACACCAACCAACGCAACCAACCAACCAACCACTACAC	288
Oy	543	accgattcccaactctcgggttaacatccacgcgtctctaataacatcatgttcaaacctgacgg	602
Db	289	CAACGACCAACAACCAACCAACCAACGACCACTTCAACAAGCAAGCACTCTAGACAAACCACTACGA	348
Oy	603	ttgtctgtagaacctcaacgcgtacatcttgatctcaataactcttaactctgtcttgacacaagaact	662
Db	349	CTACGACCAACTACTCTCTACACCACTACCAACCAACGACCAACCGGAACCAACCAACTACAA	408
Oy	663	gaacgaaaaagaatccaagaacctctacgacaaccaagttccaattctcgttatcctgaaaga	722
Db	409	CAACGACCAACAACCAACCAACCAACTTACAAAGCAACAAGCAAGCAAGCACTAGACAACTACAAAGA	468
Oy	723	cttctcgtggtgaactaactcgtcgagtagacacaacccgtac	759
Db	469	CTACGACCAACTCTTACGACGACATCTACCAACCACTAC	505

RESULT 10
BM278657

LOCUS	BM278657	569 bp	mRNA	linear	EST 20-DEC-2003
DEFINITION	As_tg2-65E06-SKL Ascaris suum adult male testis germinal zone from				
	Aln Scott Ascaris suum cDNA clone As_tg2-65E06 5', mRNA sequence.				

ACCESSION	BM278657
VERSION	BM278657.1
	GI:17971915

KEYWORDS	EST.
SOURCE	pig roundworm.

ORGANISM *Ascaris suum*
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

! Ascariidae: Ascaris.
1 (bases 1 to 569)
Blaxter, M.L., Parkinson, J., Whitton, C., Daub, J., Gulliano, D., Hall,
N., Quayle, M. and Barrrell, B.
Edinburgh University/Sanger Centre Nematode EST project
Unpublished (2000)
Contact: Blaxter M.L.

Institute of Cell, Animal and Population Biology
University of Edinburgh
Astonworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared by Michelle Lizoette-Waniewski for Alan
Scott, Johns Hopkins University Medical School, Baltimore.
Sequencing was performed by the Pathogen Sequencing Unit, Sanger
Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
sequence contained a Polya tail (trimmed)
PCR primers
FORWARD: T3
BACKWARD: T7PL
Plate: 65 Row: E Column: 06
Seq primer: SKPL
High quality sequence stop: 518.

FEATURES	location/Qualifiers
source	1. .569

```

/organism="Ascaris suum"
/db.xref="taxon:6253"
/clone="As_tgz_65E06"
/clone_lib="Ascaris suum adult male testis germinal zone
from Alan Scott"

```

```

BASE COUNT      213 a      164 c      69 g      123 t
ORIGIN
/sex="Male"
/dev.stage="Adult"
/notice="Vector: lambda zap II; Site.1: EcoRI; Site.2: XhoI"
library was made from dissected testis germinal zone from
adult male Ascaris sumu collected from abattoirs.
Constructed by Michelle Licoite-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."

```

Query Match	3.8%	Score 50.4	DB 10	length 569
Best Local Similarity	48.9%	Pred. No. 0.035		
Matches 135, Conservative	0	Mismatches 141	Indels 0	Gaps 0

Accession	Sequence	Length
OY	actctcagatgatcaaacctctctgtaacatacgaatcgatctcttggttacatccca	487
Db	119 ACAGTAATTTCAACAACGCAATTTCAACAGCAACTCAGCAACTTCATATACAACTT	178
OY	488 acaatcgctgaacaactccaaaatcacaatcaagcgctctgatgcagcagaaccca	547
Db	179 GCAACTCTCTGACAACTTCATATACCACTCAGCAACTCTCTGAGCAACTCAGTACA	238

QY	608	gtagacatcacccgtatcatcgtgataacttcaatcgtctgacgaagaactgacg	667
Db	239	ATTCACACATTAATCTTACACACCTTTAGCAGACGACCACTTTCAMAGCAGTAACTTCAACA	358

Qy	668	aaaagaataccaagaacctgtacgacaacccgtcca	703
Db	359	ACAGTAACCTCAGCACTTCAACAGCAACACTTCA	394

DEFINITION	LOCUS	RESULT 11
f551b1.1	641 bp	BM181884
Sugano SJD adult male	mRNA	BM181884
Danio rerio	linear	BM181884
EST 11-DEC-2001	clone	BM181884
5412044	5'	BM181884

ACCESSION	BM181884
VERSION	BM181884.1
KEYWORDS	GI:17512842 EST.

ORGANISM Danto rer.

REFERENCE
Clark, M., Johnson, S.-L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
1 (bases 1 to 641)
Cyprinidae: Danio.
Euarxota, Metazoa; Chordata; Crinata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
Cyprinidae: Danio.

K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swailer, T., Gibbons, M., Pape, D., Harvey, N., Schurt, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

TITLE Mashu zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: zbraflsh@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clome
distribution information can be found through the I.M.A.G.E.
Consortium/LINL, send email to: info@image.lnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 527.

FEATURES
source

```

/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5412044"
/clone_lib="Sugano SUD adult male"
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notice="vector: pME18S-FL3; site_1: DraIII (CACCATGTG); site_2: DraIII (CAGCTGTGTG); 1st strand cDNA was primed with an oligo(dT) primer [ATGTGGCCCTTTTTTTTTTTTTTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTGGCCATGTGc], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CAGCTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed and donated by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCTGCTCTTAAGCTGCG and 3' end primer GCACCTGCAGCTCGAGACCA. "
```

Query Match	3.8%	Score	50	DB	10	Length	641
Best Local Similarity	46.3%	Pred. No.	0.047				
Matches	201	Conservative	0	Mismatches	230	Indels	3
						Gaps	1

Oy	226	tacactctatgaagaaacctccacccttccttgatcgcgatcccgaatacttc	285
Dd	175	TACATCAACTTCTCCACAACAACAACTTTCTCCACAACAACAATTCTCCAGGTACATC	233
Oy	286	aactccatctctctgaaacaatgaattacacacatcaactcatgysaaaaca---ttct	342
Dd	235	AATCTTAATCAACAATAAATATCTTCCTGCGTGAATCAATCTTAACAACAATAATCTTGCG	293

Oy 343 ggttgaagatctctgaactacggtgaatcatctgaactctgcagagacactcagaa 402
 Db 295 AGCTCAACCTTCTCCAGCAACACTCACTTCCTCCACTACATCACTTCCTCAACAC 355

Db 355 AACACCTTCAGCTACATCAACTTCTCCACAACAACAACCTTCTCCAGCTACATCAAC 414

[illegible]

Db 415 TTCTCCACACACAACTTCTCCAGTCATCACTTCAACAACACAACTTCTCC 4174

DD * / 2 A0C1A0C011C00C00C00C00C011C1C0C0B01A0C00C11C1C0C0B0C0A0C0A0C0A1C 037

Oy 553 atcatgttcaaacctggaacglttgctggtaaccatcacgcgtacatcigtataaaattactic 647
| | | | |
Db 535 AACTTCTCCAGGTACATCACTTCAACAACAACAACAACCTTCTCCAGGTACATCACTTC 594
| | | | |

Qy	643	aatctgttcgaca	656
Db	595	AACACACACACAA	608

RESULT 12
A1389106

LOCUS	645 bp	mRNA	linear	EST 23-APR-2001
DEFINITION	CH20192.5	prime	CH	Drosophila melanogaster head
	melanogaster	cDNA	clone	CH20192.5, mRNA sequence.

ACCESSION	AI389106
VERSION	AI389106.2
	GI:13758704

KEYWORDS	EST.
SOURCE	fruit fly.

ORGANISM *Drosophila melanogaster*
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

RESULT	13		
CNS02156			
LOCUS		827 bp	DNA
DEFINITION	CNS02156		linear
	Tetradon nigroviridis genome survey sequence PUC-Orl end of clone		GSS 12-MAY-2000
	224F10 of library G from Tetradon nigroviridis, genomic survey		
	sequence.		
ACCESSION	AL176451		
VERSION	AL176451.1	GI:7814508	
KEYWORDS	GSS: genome survey sequence.		
SOURCE	Tetradon nigroviridis.		

Oy	423	caaat	actctga	gtgtca	aacctctgtg	actata	caatcgcgtg	atctgt	ctgtaacat	482
Db	166	CGAGAACACACAA	CAAGCAACCAACAA	CAATTAACAATTAAC	GCATTAACAACATTAACA	225				
Oy	483	cacca	aatcgtctg	aatactcca	aatctacatca	agcgctgtg	tcgac	caagaa	542	
Db	226	CAACACACACAA	CAATTAACAACAA	CAATTAACAGG	RAATTAACAACATTAACA	CAACAACA	285			
Oy	543	acg	atctcca	atctgg	taacatcca	agcttcta	atacca	atgttca	acttga	cg9
Db	286	TAAACACACAA	CAATTAACGCS	TAAACAACAATTAAC	AGCRAATTAACAAC	CAATTAACAACA	CAAA	345		
Oy	603	ttgtc	gtgac	cttacc	gtctacatctg	atcca	atctta	atctgttgc	aaga	act
Db	346	CAACATTAACA	CAACATTAAC	CCSTAAACCAATTAAC	AGCAATTAACA	CAACATTAACA	405			
Oy	663	gaac	gaaaaa	aaatcca	aagccctg	tacga	caac	caag	ttccat	tctgtga
Db	406	CAATTAACG	TGATTAACAAC	CAATTAACA	CAATTAAC	CAACATTAAC	AGCATTAACA	CA	465	
Oy	723	cttct	gggtg	tacta	ctctgc	gaag	tacga	aaacg	tacta	ctgtga
Db	466	TGACAG	CGATG	RCACAC	RACRCAACA	CAACAAC	CAACAAATTAAC	AGCATTAAC	GCRA	TAA
Oy	783	gaaca	aatag	tgtac	gtcga	taaaatg	808			
Db	526	CAACATTAAC	AGTGATTAACA	CAACAATTAAC	CAATTAAC	CAATTAAC	CAATTAAC	CAATTAAC	CAATTAAC	CAATTAAC

RESULT 14

LOCUS	BP050030	481 bp	mRNA	linear	EST 16-OCT-2000
DEFINITION	As_tgz_21C12-SKPL Ascaris suum adult male testis germinal zone from Alan Scott Ascaris suum cDNA clone As_tgz_21C12 5', mRNA sequence.				
ACCESSION	BF050030				
VERSION	BF050030.1 GI:10803926				
KEYWORDS	EST.				
SOURCE	pig roundworm.				
ORGANISM	Ascaris suum				
REFERENCE	Eukaryote; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea; Ascaridae; Ascaris.				
AUTHORS	1 (bases 1 to 481) Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guiliano,D., Hall,N., Quayle,M. and Barrell,B				
TITLE	Edinburgh University/Sanger Centre Nematode EST Project				
JOURNAL	Unpublished (2000)				
COMMENT	Contact: Blaxter ML Institute of Cell, Animal and Population Biology University of Edinburgh Aspworth labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JF, UK. Tel: +44 131 650 6760 Fax: +44 131 670 5450 Email: mark.blaxter@ed.ac.uk The library was prepared by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore. Sequencing was performed by Claire Whitton ICAPB, Edinburgh PCR Primers FORWARD: T3 BACKWARD: T7/PL Plate: 21 row: C column: 12 Seq primer: SKPL High quality sequence stop: 456. Location/Qualifiers 1..481 /organism="Ascaris suum" /db_xref="taxon:6253" /clone="As_tgz_21C12" /clone_1id="Ascaris suum adult male testis germinal zone from Alan Scott" /sex="Male" /dex_string="Adult" /note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI; Library was made from dissected testis germinal zone from adult male Ascaris suum collected from abattoirs. Constructed by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore, MD."				
BASE COUNT	188 a 149 c 49 g 95 t				
ORIGIN					
Query Match	3.7%; Score 49; DB 10; Length 481;				
Best Local Similarity	49.1%; Pred. No. 0.075;				
Matches 130; Conservative	0; Mismatches 135; Indels 0; Gaps 0;				
439	atcaacatcttgatcatatcaatctgcgcggatcttcgtttacatcatccaacatcgctcg	498			
QY					
Db	104 AACCAACACCAATTTCAACAGCAACCTCAGCAACTTCAATTAACAACTTAGCAACTCTCG	163			
QY	aataactccaatatcatatcaacagcgcgctgtatcgacacgaanaacgcatccatctg	558			
Db	164 AACCACTTCATTAACAACTCTCAGCAACGCTCTGACACAACTCAGTACACTTGAAAC	223			
QY	559 ggttaacatccagcgtcttaataaacatcatgttcaaatgtagcggttcgttgaaactaac	618			
Db	224 CTCAAACAACCTCATTAACAACTCTCAACATCTCAGCAACAACTTCAGCAATCTCAACAT	283			
QY	cgttaacatctgatacaaatcttaaatgtttgcacaagaactgaacgaanaagaatc	678			
Db	284 AACCTCAACAACTTAGCAGCAGCAACTTCAACAGCACTTAACCTTCAACAACTGTAAC	343			
QY	679 aaagacctgtaagacaaccagtcaca	703			

Db	344	AGCAACTTCAACAGCAACAATTCA	368
RESULT	15		
LOCUS	A0060230		
DEFINITION	A0060230 Dictyostelium discoidium SL (H.Urushihara) Dictyostelium		
ACCESSION	A0060230		
VERSION	A0060230.1		
KEYWORDS	EST.		
SOURCE	Dictyostelium discoidium.		
ORGANISM	Dictyostelium discoidium.		
REFERENCE	Eukaryota: Mycetozoa: Dictyostelida: Dictyostelium.		
AUTHORS	1 (bases 1 to 628) Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitta,B.N., Pl,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,T., Ochiai,H. and Tanaka,Y. Unpublished (1998) Developmental cDNA in Dictyostelium discoidium		
TITLE	Dictyostelium discoidium		
JOURNAL	Contact: Hideko Urushihara		
COMMENT	Institute of Biological Sciences University of Tsukuba 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan Email: d402huesakura.cc.tsukuba.ac.jp PROJECT = Dictyostelium discoidium cDNA project in Japan. Location/Qualifiers 1..628 /organism="Dictyostelium discoidium" /strain="AX4" /db_xref="taxon:44689" /clone="SLA628" /clone_id="Dictyostelium discoidium SL (H.Urushihara)" /dev_stage="slug"		
BASE COUNT	270 a 160 c 56 g 142 t		
ORIGIN			
Query Match	3.7%; Score 48.8; DB 9; Length 628;		
Best Local Similarity	48.6%; Pred. No. 0.094;		
Matches 134; Conservative 0; Mismatches 142; Indels 0; Gaps 0;			
QY	423 caaatctctgaatgatcaacatctctgacttaacatcgctgcatcttctgttaccat 482		
Db	78 CAACCCACCATATCAACCAACCAACACACACACACACACACACACACACACAA 137		
QY	483 caccacaacatgctcgaataatccaaatctacatacgaaggcgctgcatgcagaga 542		
Db	138 CAATTACACATATTAACAACAACAACACACACACACACACACACACACACAA 197		
QY	543 accgatcccaatctgggttaacatccacgcgttctaataacatcatttcaaatctgacg 602		
Db	198 CAC 257		
QY	603 ttgtctgtagacatccacgcgtcattcttgatcaaatattcaatctgttcgacaaagaact 662		
Db	258 ATTCAACACACATATTAACAACAACAACATCAACACACACACATCAACCA 317		
QY	663 gaacgaaaaagaatacagaagcctgtacgaacaa 698		
Db	318 CACACACACACTTAATCAATATGTTTATTCAACACACAA 353		

Search completed: September 16, 2002, 19:28:15
Job time: 8372 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 21:43:19 : Search time 631.73 Seconds
(without alignments)
3603.800 Million cell updates/sec

Title: US-09-611-419a-5
Perfect score: 1326
Sequence: 1 gattcgaacagatgacctc.....aacgtccgtcgtgaagattc 1326

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1326	100.0	22	AAA54484	Botulinum toxin hea
2	1310	98.8	22	AAA54483	Botulinum toxin hea
3	1308.2	98.7	22	AAA54482	Botulinum toxin hea
4	1302.2	98.2	17	AAAT9245	Type A neurotoxin
5	1302.2	98.2	1330	AAV30571	Clostridium botuli
6	1302.2	98.2	1338	AAZ87212	DNA encoding synth
7	1302.2	98.2	1351	AAV30576	Clostridium botuli
8	1302.2	98.2	1402	AAAT9246	Type A neurotoxin
9	1302.2	98.2	1402	AAV30572	Clostridium botuli

10	1247.8	94.1	1332	22	AAA54588	Sequence encoding
11	727.2	54.8	1546	19	AAV30575	Clostridium botuli
12	727.2	54.8	4835	21	AAAC64582	BONT/A neurotoxin
13	726	54.8	1317	21	AAZ87220	DNA encoding nativ
14	726	54.8	2532	21	AAZ87218	DNA encoding nativ
15	685.2	51.7	3891	17	AAAT9244	C. botulinum type
16	685.2	51.7	702	21	AAZ87221	DNA encoding BONTA
17	619	46.7	621	21	AAZ87222	DNA encoding BONTA
18	422	31.8	1313	18	AAAT48101	Immunogenic type F
19	422	31.8	1314	22	AAA54499	Botulinum toxin hea
20	400.6	30.2	1317	21	AAZ87216	DNA encoding synth
21	398.4	30.0	1317	22	AAA54490	Botulinum toxin hea
22	371.4	28.0	1917	21	AAZ87219	DNA encoding nativ
23	289.4	21.8	1347	21	AAZ87217	DNA encoding synth
24	289.4	21.8	1368	22	AAA54491	Botulinum toxin hea
25	264.8	20.0	1347	22	AAA54589	Sequence encoding
26	263	19.8	1341	21	AAZ87213	DNA encoding synth
27	263	19.8	1341	22	AAA54485	Botulinum toxin hea
28	261.8	19.7	1278	22	AAA54489	Botulinum toxin hea
29	261.8	19.7	1400	21	AAZ87215	DNA encoding synth
30	261.8	19.7	1400	22	AAA54488	Botulinum toxin hea
31	227	17.1	1293	18	AAAT48100	Immunogenic type F
32	202.8	15.3	1374	22	AAA54487	Botulinum toxin hea
33	193.2	14.6	1472	19	AAV30585	Clostridium botuli
34	193.2	14.6	4017	22	AAZ87219	C botulinum BONT/E
35	190	14.3	1463	19	AAV30584	Clostridium botuli
36	180.8	13.6	1460	19	AAV30593	Clostridium botuli
37	168.8	12.7	1371	21	AAZ87214	DNA encoding synth
38	168.8	12.7	1371	22	AAA54486	Botulinum toxin hea
39	166.4	12.5	1359	12	AAO12121	Synthetic tetanus
40	166.4	12.5	3712	16	AAO97490	Plasmid pHTPA1. N
41	166.4	12.5	3754	15	AAO57880	Intermediate plasm
42	166.4	12.5	3754	16	AAO55420	Plasmid pTECH1. S
43	166.4	12.5	3769	15	AAO57881	Intermediate plasm
44	166.4	12.5	3769	16	AAO97492	Plasmid pTECH2. N
45	166.4	12.5	4366	16	AAO85424	Plasmid pTECH3-P28

ALIGNMENTS

RESULT 1	AAA54484	standard; DNA; 1326 .bp.
ID	AAA54484	
AC	AAA54484;	
XX		
DT	11-APR-2001	(first entry)
DE	Botulinum toxin heavy chain C-terminal coding sequence (serotype A).	
XX		
KW	Botulinum; toxin; neurotoxin; heavy chain; recombinant expression;	
KW	recombinant vector; antigen; Immune response; vaccine; Bacterium;	
KW	Infection; ds.	
XX		
OS	Synthetic.	
OS	Clostridium botulinum.	
XX		
EH	Key	Location/Qualifiers
FT	CDS	13..1320
FT		/*tag= a
FT		/product= H_C peptide fragment
XX		
PN	WO200067700-A2.	
XX		
PD	16-NOV-2000.	
XX		
PF	12-MAY-2000; 2000WO-US12890.	
XX		
PR	12-MAY-1999; 99US-0133865.	
PR	12-MAY-1999; 99US-0133866.	
PR	12-MAY-1999; 99US-0133867.	
PR	12-MAY-1999; 99US-0133868.	

PR 12-MAY-1999; 99US-0133869.
 PR 12-MAY-1999; 99US-0133873.
 PR 29-JUL-1999; 99US-0146192.
 XX
 PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
 XX
 PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
 XX
 DR MPI: 2001-016048/02.
 DR P-FSDB: AAB04090.
 XX
 PT New nucleic acids encoding the carboxy- or amino-terminal portions of
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
 PT vaccine against botulism
 XX
 PS Disclosure: Fig 3a; 73pp; English.
 XX
 CC Botulinum neurotoxins are translated as a single 150 kDa polypeptide
 CC chain and then posttranslationally nicked, forming a dichain
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
 CC remain linked by a disulfide bond. Nucleic acids encoding the
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
 CC expression vectors and expressed in transformed cells to produce
 CC peptide antigens useful for eliciting an immune response to give
 CC protective immunity against botulinum neurotoxin, which causes
 CC botulism. The nucleic acids are expressible in a recombinant
 CC organisms such as Escherichia coli or Pichia pastoris. The use
 CC of recombinant nucleic acids are advantageous since it eliminates
 CC the need to culture large quantities of hazardous toxin-producing
 CC bacterium. Production yield from the genetically engineered product
 CC is also high and cost of production is lower. The nucleic acids can
 CC be derived from Clostridium botulinum serotypes A-G.
 XX
 SO Sequence 1326 BP; 404 A; 336 C; 243 G; 343 T; 0 other;

Query Match 100.0%; Score 1326; DB 22; Length 1326;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaattcgaaacgatgagcctctacattcattgataatcaagaacatcatactacc 60
 DB 1 gaattcgaaacgatgagcctctacattcattgataatcaagaacatcatactacc 60
 QY 61 atctggaacctgctgtaacgaatccaatcacctgacacgtgtctcgtacgcttccaa 120
 DB 61 atctggaacctgctgtaacgaatccaatcacctgacacgtgtctcgtacgcttccaa 120
 QY 61 atccgaacctgctgtaacgaatccaatcacctgacacgtgtctcgtacgcttccaa 120
 DB 61 atccgaacctgctgtaacgaatccaatcacctgacacgtgtctcgtacgcttccaa 120
 QY 121 atcaacatcggttctaagaatttaacttgatccgatacgaagaatcagatcagctgttc 180
 DB 121 atcaacatcggttctaagaatttaacttgatccgatacgaagaatcagatcagctgttc 180
 QY 121 atcaacatcggttctaagaatttaacttgatccgatacgaagaatcagatcagctgttc 180
 DB 121 atcaacatcggttctaagaatttaacttgatccgatacgaagaatcagatcagctgttc 180
 QY 181 aatcgtgaatcttccaaaatcgaaatgtaacccgaagaatcgtatcgtatacaactctatg 240
 DB 181 aatcgtgaatcttccaaaatcgaaatgtaacccgaagaatcgtatcgtatacaactctatg 240
 QY 181 aatcgtgaatcttccaaaatcgaaatgtaacccgaagaatcgtatcgtatacaactctatg 240
 DB 181 aatcgtgaatcttccaaaatcgaaatgtaacccgaagaatcgtatcgtatacaactctatg 240
 QY 241 taagaaatcttccactctctctgtagatcgatccgatacgaagaatcactcaactcatct 300
 DB 241 taagaaatcttccactctctctgtagatcgatccgatacgaagaatcactcaactcatct 300
 QY 241 taagaaatcttccactctctctgtagatcgatccgatacgaagaatcactcaactcatct 300
 DB 241 taagaaatcttccactctctctgtagatcgatccgatacgaagaatcactcaactcatct 300
 QY 301 ctgaacatgaaatacacatcattcaactgatacgaagaatcattcgttggaagaatct 360
 DB 301 ctgaacatgaaatacacatcattcaactgatacgaagaatcattcgttggaagaatct 360
 QY 301 ctgaacatgaaatacacatcattcaactgatacgaagaatcattcgttggaagaatct 360
 DB 301 ctgaacatgaaatacacatcattcaactgatacgaagaatcattcgttggaagaatct 360
 QY 361 ctgaactacggtgaaatcattcgtgactctgcaagacactcagaagaatcaacaagcgtgt 420
 DB 361 ctgaactacggtgaaatcattcgtgactctgcaagacactcagaagaatcaacaagcgtgt 420
 QY 421 gtattcaaatctctcagatgatacaacatctctgactacatacaatcgtgtcttctgt 480
 DB 421 gtattcaaatctctcagatgatacaacatctctgactacatacaatcgtgtcttctgt 480

QY 481 accatcacaacaatctgctgaaataaccccaaatctacatacaacgagcgtctgatacgc 540
 DB 481 accatcacaacaatctgctgaaataaccccaaatctacatacaacgagcgtctgatacgc 540
 QY 541 cagaacccgatactccatctggttaacatccacgcttctaataacatcgttcaacgc 600
 DB 541 cagaacccgatactccatctggttaacatccacgcttctaataacatcgttcaacgc 600
 QY 541 cagaacccgatactccatctggttaacatccacgcttctaataacatcgttcaacgc 600
 DB 541 cagaacccgatactccatctggttaacatccacgcttctaataacatcgttcaacgc 600
 QY 601 gacggttcgtgacacatccgctacatcgtatgataaataactcaatctgttcgacaa 660
 DB 601 gacggttcgtgacacatccgctacatcgtatgataaataactcaatctgttcgacaa 660
 QY 601 gacggttcgtgacacatccgctacatcgtatgataaataactcaatctgttcgacaa 660
 DB 601 gacggttcgtgacacatccgctacatcgtatgataaataactcaatctgttcgacaa 660
 QY 661 gaactgaacgaaaaaaagaatcaagaacctgtacgacacacagtcacatctgtatccgt 720
 DB 661 gaactgaacgaaaaaaagaatcaagaacctgtacgacacacagtcacatctgtatccgt 720
 QY 721 aaagactctgggtgtaacacccgctgacgaagaacgcttactaactggtatctgtac 780
 DB 721 aaagactctgggtgtaacacccgctgacgaagaacgcttactaactggtatctgtac 780
 QY 781 gatccgaacaatacgttgaacgaagaatgtagatccgctggttatacgtacgtgaa 840
 DB 781 gatccgaacaatacgttgaacgaagaatgtagatccgctggttatacgtacgtgaa 840
 QY 841 ggtccgctggttctgtttagactacaacatctacttcaactcttccctgtacgtgt 900
 DB 841 ggtccgctggttctgtttagactacaacatctacttcaactcttccctgtacgtgt 900
 QY 901 accaaatcatcatcaagaataacgctgtgtgaacaagaacatacgttcgacaacat 960
 DB 901 accaaatcatcatcaagaataacgctgtgtgaacaagaacatacgttcgacaacat 960
 QY 961 gatcgtgatacaatcaatgtttagtgaagaacaagaataccgttcgtacccaatgt 1020
 DB 961 gatcgtgatacaatcaatgtttagtgaagaacaagaataccgttcgtacccaatgt 1020
 QY 1021 tctcagcgtggtgtgagaagaatctgtctgtctgtgaatcccgagcgtgtgatacgt 1080
 DB 1021 tctcagcgtggtgtgagaagaatctgtctgtctgtgaatcccgagcgtgtgatacgt 1080
 QY 1081 tctcagcgtggtgtgagaagaatctgtctgtctgtgaatcccgagcgtgtgatacgt 1140
 DB 1081 tctcagcgtggtgtgagaagaatctgtctgtctgtgaatcccgagcgtgtgatacgt 1140
 QY 1141 aatctgcaagacaacaatgtaacgatalcgtgttcaatcgtgttccacagttcaaat 1200
 DB 1141 aatctgcaagacaacaatgtaacgatalcgtgttcaatcgtgttccacagttcaaat 1200
 QY 1201 atcgtcaaatggtgtgtcccaactggtatcaatcgtatcgaatcgaatcgttccctgcact 1260
 DB 1201 atcgtcaaatggtgtgtcccaactggtatcaatcgtatcgaatcgaatcgttccctgcact 1260
 QY 1261 ctgggtgtctcttggtgagttacatcccggtgtgtgagcgtgttggtgagtcgcgttaa 1320
 DB 1261 ctgggtgtctcttggtgagttacatcccggtgtgtgagcgtgttggtgagtcgcgttaa 1320
 QY 1321 gaattc 1326
 DB 1321 gaattc 1326

RESULT 2
 AAA54483
 ID AAA54483 standard; DNA; 1323 BP.
 XX
 AC AAA54483;
 XX
 DT 11-APR-2001 (first entry)
 XX
 DE Botulinum toxin heavy chain C-terminal coding sequence (serotype A).
 XX Botulinum toxin; neurotoxin; heavy chain; recombinant expression;
 KW recombinant vector; antigen; immune response; vaccine; bacterium;
 KW

KM infection: ds.
 XX Synthetic.
 OS Clostridium botulinum.
 XX
 FH Key Location/Qualifiers
 FT 13..1317
 FT CDS /*tag- a
 /product= H_C peptide fragment
 XX
 PN MO200067700-A2.
 XX
 PD 16-NOV-2000.
 XX
 PF 12-MAY-2000; 2000MO-US12890.
 XX
 PR 12-MAY-1999; 99US-0133865.
 PR 12-MAY-1999; 99US-0133866.
 PR 12-MAY-1999; 99US-0133867.
 PR 12-MAY-1999; 99US-0133868.
 PR 12-MAY-1999; 99US-0133869.
 PR 12-MAY-1999; 99US-0133873.
 PR 29-JUL-1999; 99US-0146192.
 XX
 PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
 XX
 PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
 XX
 DR WPI: 2001-016048/02.
 DR P-PSDB; AAB04089.
 XX
 XX
 PT New nucleic acids encoding the carboxy- or amino-terminal portions of
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
 PT vaccine against botulism
 XX
 PS Disclosure; Fig 2a; 73pp; English.
 XX
 XX Botulism neurotoxins are translated as a single 150 kDa polypeptide
 CC chain and then posttranslationally nicked, forming a dichain
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
 CC remain linked by a disulfide bond. Nucleic acids encoding the
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
 CC expression vectors and expressed in transformed cells to produce
 CC peptide antigens useful for eliciting an immune response to give
 CC protective immunity against botulinum neurotoxin, which causes
 CC botulism. The nucleic acids are expressible in a recombinant
 CC organisms such as Escherichia coli or Pichia pastoris. The use
 CC of recombinant nucleic acids are advantageous since it eliminates
 CC the need to culture large quantities of hazardous toxin-producing
 CC bacterium. Production yield from the genetically engineered product
 CC is also high and cost of production is lower. The nucleic acids can
 CC be derived from Clostridium botulinum serotypes A-G.
 XX
 XX Sequence 1323 BP; 404 A; 334 C; 242 G; 343 T; 0 other;
 SQ

Query Match 98.8%; Score 1310; DB 22; Length 1323;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1323; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 gaattcgaaacgattgctctaccctacatgaatcaatcaagaacatcatcaatcctcc 60
 Db 1 gaattcgaaacgattgctctaccctacatgaatcaatcaagaacatcatcaatcctcc 57
 QY 61 atcctgaacctgctgaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 120
 Db 58 atcctgaacctgctgaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 117
 QY 121 atcaaatcggtcttaagttaactgcatcgatcgatcgatcgatcgatcgatcgatcgatc 180
 Db 118 atcaaatcggtcttaagttaactgcatcgatcgatcgatcgatcgatcgatcgatcgatc 177

QY 181 aatctgaatctcccaaatcgaaagtatactctgaagaagtctatcgatatacaactctatg 240
 Db 178 aatctgaatctcccaaatcgaaagtatactctgaagaagtctatcgatatacaactctatg 237
 QY 241 taagaaactctcccaactctctctgtgataccgatacccgaaatactcaactcctctct 300
 Db 238 taagaaactctcccaactctctctgtgataccgatacccgaaatactcaactcctctct 297
 QY 301 ctgaacaaatgaatcacatcatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 360
 Db 298 ctgaacaaatgaatcacatcatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 357
 QY 361 ctgaactacggtgaatcatcatctgactctgacgacacacacacacacacacacacacac 420
 Db 358 ctgaactacggtgaatcatcatctgactctgacgacacacacacacacacacacacacac 417
 QY 421 gtattcaaatctctcagaatcaatcaatctctgatacatcaatcaatcaatcaatcaat 480
 Db 418 gtattcaaatctctcagaatcaatcaatctctgatacatcaatcaatcaatcaatcaat 477
 QY 481 accatcccaacaaatcgctgaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 540
 Db 478 accatcccaacaaatcgctgaatcaatcaatcaatcaatcaatcaatcaatcaatcaatcaat 537
 QY 541 cagaacccgatactcccaatctggtatacaccacacacacacacacacacacacacacac 600
 Db 538 cagaacccgatactcccaatctggtatacaccacacacacacacacacacacacacacac 597
 QY 601 gacggttgctgtaacatcaccacacacacacacacacacacacacacacacacacacac 660
 Db 598 gacggttgctgtaacatcaccacacacacacacacacacacacacacacacacacacac 657
 QY 661 gaactgaacgaaagaagaatcaagaactctgacgacacacacacacacacacacacacac 720
 Db 658 gaactgaacgaaagaagaatcaagaactctgacgacacacacacacacacacacacacac 717
 QY 721 aaagactctggtgactactctgactgacgacacacacacacacacacacacacacacacac 780
 Db 718 aaagactctggtgactactctgactgacgacacacacacacacacacacacacacacacac 777
 QY 781 gatccgaacaaatcagttgacgtcaacaatgtagtalcgoggttaccatgtaacctgaa 840
 Db 778 gatccgaacaaatcagttgacgtcaacaatgtagtalcgoggttaccatgtaacctgaa 837
 QY 841 ggtccggtggttcgtgatactcaacacacacacacacacacacacacacacacacacacac 900
 Db 838 ggtccggtggttcgtgatactcaacacacacacacacacacacacacacacacacacacac 897
 QY 901 accaatcatcatcaagaagaatcagcgtctgtaacaggaacacacacacacacacacacac 960
 Db 898 accaatcatcatcaagaagaatcagcgtctgtaacaggaacacacacacacacacacacac 957
 QY 961 gatcgtatcatcaatcagttgtagttaagaacaaatcagcgtctggtaccacacacac 1020
 Db 958 gatcgtatcatcaatcagttgtagttaagaacaaatcagcgtctggtaccacacacac 1017
 QY 1021 tctcaaggtggtgtagaagaac 1080
 Db 1018 tctcaaggtggtgtagaagaac 1077
 QY 1081 tctcaaggtggtgtagaagaac 1140
 Db 1078 tctcaaggtggtgtagaagaac 1137
 QY 1141 aatctgacgacaaac 1200
 Db 1138 aatctgacgacaaac 1197
 QY 1201 atcgttaaacgtgtgcttccac 1260
 Db 1198 atcgttaaacgtgtgcttccac 1257
 QY 1261 ctgggttgctcttgaggatcatcccggttgatgacggttgagggtgacgctcgctgtaa 1320

```

Db 1258 ctggcttgccttggagtcacccggttgagtcagcgttggggtgaacgtccgttaa 1317
Qy 1321 gaattc 1326
   |||||
Db 1318 gaattc 1323

RESULT: 3
AAAS4482
ID AAAS4482 standard; DNA; 1332 BP.
XX
AC AAAS4482;
DT 11-APR-2001 (first entry)
DE Botulinum toxin heavy chain C-terminal coding sequence (serotype A).
KM Botulinum toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; Dactylum;
XX Infection; ds.
OS Synthetic.
XX Clostridium botulinum.
FH Key
FT CDS Location/Qualifiers
   /*tag= a
   /product= H_C peptide fragment
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PE 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JU, Lapenotiere H;
XX
DR WPI: 2001-016048/02.
DR P-PSDB: AAB04088.
XX
XX New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulinum
XX
XX Claim 2; Fig 1a; 73pp; English.
XX
XX Botulinum neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.

```

```

XX SQ Sequence 1332 BP; 404 A; 337 C; 245 G; 346 T; 0 other;
Query Match          98.7%; Score 1308.2; DB 22; Length 1332;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1310; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 14 tggcctcactcctcactgaatacatcaagaacatcatcaatacctccatcgtgaacctgc 73
   || |||||
Db 20 tgcgtctcactcctcactgaatacatcaagaacatcatcaatacctccatcgtgaacctgc 79
Qy 74 gctacgaatccaatacctcgtgacccgtgctcgtctacgcttccaaatccaacgcggtt 133
   |||||
Db 80 gctacgaatccaatacctcgtgacccgtgctcgtctacgcttccaaatccaacgcggtt 139
Qy 134 ctaaggttaactcgcgtccgatccgcgaagaatccaatccaatcgcgttccaatcgcgaatct 193
   |||||
Db 140 ctaaggttaactcgcgtccgatccgcgaagaatccaatccaatcgcgttccaatcgcgaatct 199
Qy 194 ccaaatcgaagtatcctcgaagaatgctatcgtatcaactctatgtacgaagaactctc 253
   |||||
Db 200 ccaaatcgaagtatcctcgaagaatgctatcgtatcaactctatgtacgaagaactctc 259
Qy 254 ccactcctctcgtgacgtcgtatcccgaaatacttaactccaactcctcgtgaacaatgaa 313
   |||||
Db 260 ccactcctctcgtgacgtcgtatcccgaaatacttaactccaactcctcgtgaacaatgaa 319
Qy 314 acacatcatcaacgcgtgacgaagaacaaatcctcgtctggaagaatctcgtgaactacggtg 373
   |||||
Db 320 acacatcatcaacgcgtgacgaagaacaaatcctcgtctggaagaatctcgtgaactacggtg 379
Qy 374 aaatcatctggaactctcgaagacactcaggaatccaacagcgtgtgtatccaataact 433
   |||||
Db 380 aaatcatctggaactctcgaagacactcaggaatccaacagcgtgtgtatccaataact 439
Qy 434 ctgagatgatacaactcctcgtactatcaatcgcgtcgtgactcgttcaacacacaca 493
   |||||
Db 440 ctgagatgatacaactcctcgtactatcaatcgcgtcgtgactcgttcaacacacaca 499
Qy 494 atcgtctgaataactccaacaaatctacatcaacgcgcgtctgatacgaccgaagaacgatct 553
   |||||
Db 500 atcgtctgaataactccaacaaatctacatcaacgcgcgtctgatacgaccgaagaacgatct 559
Qy 554 ccaatctgggtatacatccaacgcgtcttcaataacataatgttcaactcgtgacggttgcgtg 613
   |||||
Db 560 ccaatctgggtatacatccaacgcgtcttcaataacataatgttcaactcgtgacggttgcgtg 619
Qy 614 acactcaccgctacatctcgtatccaataacttcaatcgtctgcgaagaacgtgaacga 673
   |||||
Db 620 acactcaccgctacatctcgtatccaataacttcaatcgtctgcgaagaacgtgaacga 679
Qy 674 aagaatccaagaacactgtacgaacaacccagttccaattctgttactcgtgaagaactcttg 733
   |||||
Db 680 aagaatccaagaacactgtacgaacaacccagttccaattctgttactcgtgaagaactcttg 739
Qy 734 gtgactaccctcagtcagcaacaacccgtactacatcgtcgaactctgtacgactccgaacaat 793
   |||||
Db 740 gtgactaccctcagtcagcaacaacccgtactacatcgtcgaactctgtacgactccgaacaat 799
Qy 794 agcttgagctcaacaatgtgattatccgggttacaatgtacgtgaagaagttccgctgggtt 853
   |||||
Db 800 agcttgagctcaacaatgtgattatccgggttacaatgtacgtgaagaagttccgctgggtt 859
Qy 854 ctgtatgactccaacatctacccgtgaactctcctcgtacgctgtgtacgaatccaatcca 913
   |||||
Db 860 ctgtatgactccaacatctacccgtgaactctcctcgtacgctgtgtacgaatccaatcca 919
Qy 914 tcaagaatactcgtctgttacaagaagacatcgttgcgaacaatgtatcgtatcga 973
   |||||
Db 920 tcaagaatactcgtctgttacaagaagacatcgttgcgaacaatgtatcgtatcga 979
Qy 974 tcaatgtgtgtagttaagaacaagaatactcgtctgtcaccaaatgcttctcagctgctg 1033

```

```

|||||
Db 980 tcaatgtgttagtaagaacaagaataccgtctgcgtacccaatgcttccagctggtg 1039
QY 1034 tagaagaatctgtctgtctgtctgtgaataccggagcgttggtaactgtctgaagtgtg 1093
Db 1040 tagaagaatctgtctgtctgtctgtgaataccggagcgttggtaactgtctgaagtgtg 1099
QY 1094 taagaatccaagaacagcagggtatccaataaataatgaatgtcagagaca 1153
Db 1100 taagaatccaagaacagcagggtatccaataaataatgaatgtcagagaca 1159
QY 1154 acaatggtaacagatcgtgttcacgttccacccagttcaacaatacgtctaactgtg 1213
Db 1160 acaatggtaacagatcgtgttcacgttccacccagttcaacaatacgtctaactgtg 1219
QY 1214 ttgtctcaactgtgtacaatcgttaagatcgagcgttccctcgcacatctgtgtgtctct 1273
Db 1220 ttgtctcaactgtgtacaatcgttaagatcgagcgttccctcgcacatctgtgtgtctct 1279
QY 1274 ggaggttcaccccggttcgtagcaggttgggtgtagacgttcgcgttaagaattc 1326
Db 1280 ggaggttcaccccggttcgtagcaggttgggtgtagacgttcgcgttaagaattc 1332

```

RESULT 4

AAT29245 standard; DNA: 1330 BP.

AAT29245;

07-JUL-1996 (first entry)

Type A neurotoxin C fragment synthetic gene.

Toxin: neurotoxin; fusion protein; antitoxin; vaccine; immunogen;

Clostridium botulinum; ds.

Synthetic.

Location/Qualifiers

1..1317

/*tag- a

WO9612802-A1.

02-MAY-1996.

23-OCT-1995; 95WO-US13737.

07-JUN-1995; 95US-0480604.

24-OCT-1994; 94US-0329154.

16-MAR-1995; 95US-0405486.

14-APR-1995; 95US-0422711.

(OPHI-) OPHIDIAN PHARM INC.

Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;

Williams JA;

WPI; 1996-230603/23.

P-PSDB; AAR95008.

Fusion proteins comprising non-toxin protein and part of toxin

useful to form anti-toxins against Clostridium botulinum type A, and

C. difficile type toxins, and to treat C. difficile intoxication,

partic. diarrhoea

Example 22; Page 336-38; 434pp; English.

A synthetic gene (AAT29245) codes for the heavy chain C fragment

(AAR95008) of Clostridium botulinum type A neurotoxin (see also

AAR95010). Codon usage allowing efficient gene expression in Escherichia

coli was utilised. The gene in vector pALTERbot was used to make

CC expression constructs in which fragments of C. difficile toxin A
 CC repeat domains were expressed as genetic fusions with the C.
 CC botulin C fragment and expressed in E. coli.

XX Sequence 1330 BP; 400 A; 339 C; 246 G; 345 T; 0 other;

Query Match 98.2%; Score 1302.2; DB 17; Length 1330;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 14 tggcctcattcttcattgaatacacaagaacatcaatacctcctcctgaacctgc 73
Db 11 tgcgtctacctcattgaatacacaagaacatcaatacctcctcctgaacctgc 70
QY 74 gctacgaatccaatacctcctgaacctgcgtctcgtctacgcttccaataacatcggt 133
Db 71 gctacgaatccaatacctcctgaacctgcgtctcgtctacgcttccaataacatcggt 130
QY 134 ctaagtttaacttcgatccgatccgaagaatcagatccagctgttcaatctgtgactct 193
Db 131 ctaagtttaacttcgatccgatccgaagaatcagatccagctgttcaatctgtgactct 190
QY 194 ccaaatcgaagttaacctgaagaatgctatcgtatatacactctatgtacgaanaactct 253
Db 191 ccaaatcgaagttaacctgaagaatgctatcgtatatacactctatgtacgaanaactct 250
QY 254 ccaactcctctgtgacccgtatcccgaaataactccaactcctctctgaacatgat 313
Db 251 ccaactcctctgtgacccgtatcccgaaataactccaactcctctctgaacatgat 310
QY 314 acaccatataactcgtacgtgaanaaacttctggttgaagaatctctgaactcgtgtg 373
Db 311 acaccatataactcgtacgtgaanaaacttctggttgaagaatctctgaactcgtgtg 370
QY 374 aaatcatctggaactcgtgaagacactcaggaatacgaacagcgttcttatcaaatct 433
Db 371 aaatcatctggaactcgtgaagacactcaggaatacgaacagcgttcttatcaaatct 430
QY 434 ctacgaatgatacaatcctctgaactacatcaatcgtctgactctgttccatccaaca 493
Db 431 ctacgaatgatacaatcctctgaactacatcaatcgtctgactctgttccatccaaca 490
QY 494 atcgtctgaataactccaataactctacatcaacagcgcgtctgacccaggaactct 553
Db 491 atcgtctgaataactccaataactctacatcaacagcgcgtctgacccaggaactct 550
QY 554 ccaatctggtgaacataccaacgttcttaataacatcatgttccaactggaacgttgcgt 613
Db 551 ccaatctggtgaacataccaacgttcttaataacatcatgttccaactggaacgttgcgt 610
QY 614 acactccacgcgtacatctcgtgatccaataactccaactctgttgacaaaagaactgaa 673
Db 611 acactccacgcgtacatctcgtgatccaataactccaactctgttgacaaaagaactgaa 670
QY 674 aagaatacaagacctgtgacgaacaacagtcgaactctcgtgatacctgaagaactctgg 733
Db 671 aagaatacaagacctgtgacgaacaacagtcgaactctcgtgatacctgaagaactctgg 730
QY 734 gtgactacctgacgtacgaacaacagtcgaactctcgtgatacctgaagaactctgg 793
Db 731 gtgactacctgacgtacgaacaacagtcgaactctcgtgatacctgaagaactctgg 790
QY 794 acgttgaacgtacaacatgtagtaccgcggtttacatgtaactcgaagaagtcgcgctggt 853
Db 791 acgttgaacgtacaacatgtagtaccgcggtttacatgtaactcgaagaagtcgcgctggt 850
QY 854 ctgttatgactacaacaactacacgaactctccctctacggttgataccaatcatca 913
Db 851 ctgttatgactacaacaactacacgaactctccctctacggttgataccaatcatca 910
QY 914 tcaagaatacgcgtctgtgatacgaagaacatatacgttcgcaacaatgatcgtgtataca 973

```

Db 911 tcaagaatacgcgtctgtgtaacaagacatatcgttcgcaaatgatacgtgtatata 970
 Qy 974 tcaatgcttaagtaagaacaagaataacgctctgtgctacaaatgcttctcaaggctg 1033
 Db 971 tcaatgcttaagtaagaacaagaataacgctctgtgctacaaatgcttctcaaggctg 1030
 Qy 1034 tagaagaatcttctgtctgtgtaaatcccgagctgtgtaatactgtctcagtagtg 1093
 Db 1031 tagaagaatcttctgtctgtgtaaatcccgagctgtgtaatactgtctcagtagtg 1090
 Qy 1094 taataagaatccgaagacgaggtatcactaacaagaataagaatctgcggagaca 1153
 Db 1091 taataagaatccgaagacgaggtatcactaacaagaataagaatctgcggagaca 1150
 Qy 1154 acaatgtaacgatalcggattcaccgtgttcaccagttcaacaatacgtctaaactg 1213
 Db 1151 acaatgtaacgatalcggattcaccgtgttcaccagttcaacaatacgtctaaactg 1210
 Qy 1214 ttgcttccaaactggtacaaatcgtcagatcgaacgttctctcgcactctgtgtctct 1273
 Db 1211 ttgcttccaaactggtacaaatcgtcagatcgaacgttctctcgcactctgtgtctct 1270
 Qy 1274 gggagttcccggttgatgacggttggtggaacgttcgcgtgtgaa 1320
 Db 1271 gggagttcccggttgatgacggttggtggaacgttcgcgtgtgaa 1317

RESULT 5
 AAV30571
 ID AAV30571 standard; DNA; 1330 BP.
 AC AAV30571;

DT 07-DEC-1998 (first entry)
 DE Clostridium botulinum toxin A fragment C gene in PalterBot.
 XX Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
 KW botulism; ds.
 OS Clostridium botulinum serotype A.
 FH Key Location/Qualifiers
 FT CDS 1..1317
 FT misc_difference 1..6 /*tag= a
 FT /*tag= b
 FT /*note= "PALTER vector-derived nucleotides
 (encode Met-Ala)"

PN W09808540-A1.
 PD 05-MAR-1998.
 PE 28-AUG-1997; 97WO-US15394.
 PR 28-AUG-1996; 96US-0704159.
 PA (OPHI-) OPHIDIAN PHARM INC.
 PI Thelley BS, Williams JA;
 DR WPI: 1998-230234/20.
 DR P-PSDB: AAM68389.
 XX Host cell containing recombinant expression vector encoding
 PT Clostridium botulinum type B or E toxin - useful to treat humans
 PR and other animals at risk of intoxication with clostridial toxin
 XX
 PS Example 22; Page 262-263; 428pp; English.
 CC This is the DNA sequence of the Clostridium botulinum serotype A
 CC toxin C-fragment gene contained in plasmid PalterBot. Recombinant

CC C-fragment proteins have been produced in Escherichia coli as
 CC fusion proteins with either maltose binding protein or
 CC Clostridium difficile type A toxin (see AAM68387). The invention
 CC relates to recombinant proteins derived from C. botulinum toxins.
 CC Methods are provided which allow for the isolation of soluble
 CC recombinant proteins free of significant endotoxin contamination.
 CC Preferred hosts for production of recombinant toxin proteins are E. coli,
 CC insect cells and yeast cells. The recombinant toxin proteins are
 CC used as immunogens for the production of vaccines and antitoxins
 CC that are useful in the treatment of humans and animals at risk of
 CC intoxication with clostridial toxin.

XX Sequence 1330 BP; 400 A; 339 C; 246 G; 345 T; 0 other;

Query Match 98.2%; Score 1302.2; DB 19; Length 1330;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 14 ttgctctaccttcaactgataatatacaagaacatacataactcctcctgaactcgc 73
 Db 11 ttgctctaccttcaactgataatatacaagaacatacataactcctcctgaactcgc 70
 Qy 74 gctacgaatccaatccacccgattcgtctgtcgtcagcttccaaatccaatcgtgt 133
 Db 71 gctacgaatccaatccacccgattcgtctgtcgtcagcttccaaatccaatcgtgt 130
 Qy 134 ctaaatcttaactcgtatcccgatcgcagaagaatcagatccagctgttcaactgtgaactc 193
 Db 131 ctaaatcttaactcgtatcccgatcgcagaagaatcagatccagctgttcaactgtgaactc 190
 Qy 194 ccaaaatcgaagtatcctcgtgaagaatgctatcgtatacaactctatgtcgaactct 253
 Db 191 ccaaaatcgaagtatcctcgtgaagaatgctatcgtatacaactctatgtcgaactct 250
 Qy 254 ccaccccttcttgatccgtatcccgaaatctcaacccatctctcgtgaagaat 313
 Db 251 ccaccccttcttgatccgtatcccgaaatctcaacccatctctcgtgaagaat 310
 Qy 314 acacatcatcaactcgtacgagaagaacaattctgttggaaagtatctcgtgaactcgtg 373
 Db 311 acacatcatcaactcgtacgagaagaacaattctgttggaaagtatctcgtgaactcgtg 370
 Qy 374 aaatcatctgactctgcagacacactcaggaatacaaacaggtgtgtatcaactact 433
 Db 371 aaatcatctgactctgcagacacactcaggaatacaaacaggtgtgtatcaactact 430
 Qy 434 ctcaagtgataaacatctctgtactataatcgtcgtggtcttctgttaccatcaccaaa 493
 Db 431 ctcaagtgataaacatctctgtactataatcgtcgtggtcttctgttaccatcaccaaa 490
 Qy 494 atcgtctgataaactccaataatctacatcaacggtcgttgcagccagaaacgactc 553
 Db 491 atcgtctgataaactccaataatctacatcaacggtcgttgcagccagaaacgactc 550
 Qy 554 ccaatctgggtacatccagcgttctaataacatcagtggtcaactcgtggtgtcgtg 613
 Db 551 ccaatctgggtacatccagcgttctaataacatcagtggtcaactcgtggtgtcgtg 610
 Qy 614 acactcaccgtctacatctgtatcaaatcttaacttaacttctgttgcagaagaactcgaagaa 673
 Db 611 acactcaccgtctacatctgtatcaaatcttaacttaacttctgttgcagaagaactcgaagaa 670
 Qy 674 aagaatcaagaagcgtgtcagcaacaacagtcacatctgttgcagaaagactctcgtg 733
 Db 671 aagaatcaagaagcgtgtcagcaacaacagtcacatctgttgcagaaagactctcgtg 730
 Qy 734 gtgactacctgagtaagcaaacacgttactatactgtgaactcgtgtcgaacaaat 793
 Db 731 gtgactacctgagtaagcaaacacgttactatactgtgaactcgtgtcgaacaaat 790
 Qy 794 aggtgagctcaacaatgtaggtatccggttatactgtgaactcgaagaagtcgcggtgt 853

```

Db 791 acgttgacatcaaatgtagtgcgcggttacatgatactgaaagctccgctggtt 850
Qy 854 ctgtttgactaaccaactctacactcttccctgtaccgtggtcccaattcaca 913
Db 851 ctgtttgactaaccaactctacactcttccctgtaccgtggttacccaattcaca 910
Qy 914 tcaagaatacgcgtctgttaacaagaacatctcgttcgcgaacatctcgttataca 973
Db 911 tcaagaatacgcgtctgtgttaacaagaacatctcgttcgcgaacatctcgttataca 970
Qy 974 tcaatgttgttttaagaacaagaatacgcgtctgtgttaccatgaatcttcaagctggtg 1033
Db 971 tcaatgttgttttaagaacaagaatacgcgtctgtgttaccatgaatcttcaagctggtg 1030
Qy 1034 tagaagaatcttctgtcgtctgtgaatacccgagcgtgtgttaattctgtcgaagtggt 1093
Db 1031 tagaagaatcttctgtcgtctgtgaatacccgagcgtgtgttaattctgtcgaagtggt 1090
Qy 1094 taatgaataccaagaacagcaggtatcaactaacaatactgaatactgcagagaca 1153
Db 1091 taatgaataccaagaacagcaggtatcaactaacaatactgaatactgcagagaca 1150
Qy 1154 acaatgttaacagatcgcgttccatcgcgttccaccagttcaacaatactgcataactggtg 1213
Db 1151 acaatgttaacagatcgcgttccatcgcgttccaccagttcaacaatactgcataactggtg 1210
Qy 1214 ttgcttccaactgtgtacatcgtgtacagatcgaactctctcctgcacactctgtggtctctt 1273
Db 1211 ttgcttccaactgtgtacatcgtgtacagatcgaactctctcctgcacactctgtggtctctt 1270
Qy 1274 gggagttcatcccggttgaatgaacggttggtgggtgaacgctgcgtgtaa 1320
Db 1271 gggagttcatcccggttgaatgaacggttggtgggtgaacgctgcgtgtaa 1317

RESULT 6
AAZ87212
ID AAZ87212 standard; DNA; 1338 BP.
AC AAZ87212;
XX
XX 08-MAY-2000 (first entry)
DE DNA encoding synthetic BONT serotype A (BONTA) Hc fragment.
XX
XX Botulinum neurotoxin; heavy chain: BONT; serotype A;
KM C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KM VEE; botulinum; vaccine; diagnosis; drug screening; ds.
XX
OS Clostridium botulinum.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 9..1325
XX FT /*tag= "a
XX FT /product= "Synthetic botulinum neurotoxin serotype A
XX FT (BONTA) heavy chain C-terminal fragment (Hc)"
XX PN WO200002524-A2.
XX
XX 20-JAN-2000.
XX
XX 09-JUL-1999; 99WO-US15570.
XX
XX 10-JUL-1998; 98US-0092416.
XX 12-MAY-1999; 99US-0133870.
XX
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
XX WPI: 2000-160827/14.
XX P-PSDB; AAV77134.
DR

```

```

XX
PT Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum -
XX
XX PS Disclosure: Page 54; 54pp; English.
XX
XX The invention relates to novel vaccines that induce a protective immune
CC response against botulinum neurotoxin (BONT) serotypes A, B, C, D, E, F
CC and G (BONTA-BONTG). The vaccine of the invention is novel recombinant
CC DNA construct comprising a C-terminal heavy chain fragment (HC) from BONT
CC fragment comprising a C-terminal heavy chain fragment (HC) from BONT
CC serotypes A-G. In preferred embodiments of the invention, the vector is
CC a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
CC this vector results in the production of large amounts of a protein
CC encoded by a sequence cloned into the replicon. The constructs are used
CC to produce vaccines against botulism. The proteins can also be used as
CC diagnostic tools for the diagnosis of botulism. The transformed host
CC cells can be used to analyse the effectiveness of drugs and agents which
CC inhibit toxin effects. The vaccine currently used against botulism is
CC dangerous and expensive to produce, and contains formalin, which is very
CC painful for the recipient. Also, the vaccine is incomplete, in that only
CC 5 of the 7 serotypes are represented in the formulation. The novel
CC vaccine of overcomes these problems, as it is easily purified, and
CC available in large quantities. It is also expressed in the lymph nodes
CC for a better immune response. Sequences AAZ87212-287217 represent
CC synthetic DNA sequences encoding BONT Hc fragments used in the present
CC invention. These were optimised for codon usage for expression in yeast.
XX
XX Sequence 1338 BP; 401 A; 342 C; 249 G; 346 T; 0 other:

```

```

Query Match 98.2%; Score 1302.2; DB 21; Length 1338;
Best Local Similarity 99.8%; Pred. No. 0; Mismatches 3; Indels 0; Gaps 0;
Matches 1304; Conservative 0;
Qy 14 tggcctctactctactcgtatatacaagaacatcaatacctccatcctgaaactgc 73
Db 19 tggctgtctactctactcgtatatacaagaacatcaatacctccatcctgaaactgc 78
Qy 74 gctacgaatccaatacctgtagcactgctcgtgaagcttccaaaatacaactcgtgt 133
Db 79 gctacgaatccaatacctgtagcactgctcgtgaagcttccaaaatacaactcgtgt 138
Qy 134 ctaagtttaactcgtatccgacgaagaatacgaatcagctgttcaactcgtgaactct 193
Db 139 ctaagtttaactcgtatccgacgaagaatacgaatcagctgttcaactcgtgaactct 198
Qy 194 ccaaatcgaagttaacctgaagaatgtatctatatacaactctatgtacgaanaactct 253
Db 199 ccaaatcgaagttaacctgaagaatgtatctatatacaactctatgtacgaanaactct 258
Qy 254 ccaactcctctctgtatcgtatccgaaataactccaactctctctgtgaacatgaat 313
Db 259 ccaactcctctctgtatcgtatccgaaataactccaactctctctgtgaacatgaat 318
Qy 314 acaccatacaactcgtatggaagaatactcgtgttgaagaatctctgtgaactcgtgtg 373
Db 319 acaccatacaactcgtatggaagaatactcgtgttgaagaatctctgtgaactcgtgtg 378
Qy 374 aatcatctgtagctctgcagacactcaggaatacaaaaagcgttgttatccaataact 433
Db 379 aatcatctgtagctctgcagacactcaggaatacaaaaagcgttgttatccaataact 438
Qy 434 ctcaatgatcaaatctctgtactatcaatcgtctgtgttctgtttaccatccaaca 493
Db 439 ctcaatgatcaaatctctgtactatcaatcgtctgtgttctgtttaccatccaaca 498
Qy 494 atcgtcgaataacccaacatctatcaacagcgcgtctgtatgcagcaaacagatct 553
Db 499 atcgtcgaataacccaacatctatcaacagcgcgtctgtatgcagcaaacagatct 558
Qy 554 ccaatctgggtaacatccaacgcttctaatacatatgttcaactggaacggttgtcgtg 613

```

```
Db      559 ccaatctggtaacatccacgcttctaataacatcatgttcaaacctggacggtgttcgtg 618
Oy      614 acactaccgctacatcatgatcaaatacttcaatctgttcgacaagaactgaaagaaa 673
Db      619 acactaccgctacatcatgatcaaatacttcaatctgttcgacaagaactgaaagaaa 678
Oy      674 aagaatacaaaagcctgtacgacaacacgactccaattctgtatcctcgaagactctgg9 733
Db      679 aagaatacaaaagcctgtacgacaacacgactccaattctgtatcctcgaagactctgg9 738
Oy      734 gtagactactgtagtaccgacaacacgactccaattctgtatcctcgaagactctgg9 733
Db      739 gtagactactgtagtaccgacaacacgactccaattctgtatcctcgaagactctgg9 738
Oy      794 acgttgaactcaacaatgttaggtatccgctgtagtaccgacaagctccgctggtt 853
Db      799 acgttgaactcaacaatgttaggtatccgctgtagtaccgacaagctccgctggtt 858
Oy      854 ctgttatgactaccacaatctactgtagtaccgcttccctgtacacgctggtaaccaatcatca 913
Db      859 ctgttatgactaccacaatctactgtagtaccgcttccctgtacacgctggtaaccaatcatca 918
Oy      914 tcaagaataacgctgtctgttaacaagaacaatctgttcgacaacatgtagctgtatata 973
Db      919 tcaagaataacgctgtctgttaacaagaacaatctgttcgacaacatgtagctgtatata 978
Oy      974 tcaatctttagttaaagaacaagaatccgctgtgctacacgaatgcttccagctggtg 1033
Db      979 tcaatctttagttaaagaacaagaatccgctgtgctacacgaatgcttccagctggtg 1038
Oy      1034 tgaagaagaactctgtctgtcgtggaatcccggaacgttggtaactgtctcaggtagtg 1093
Db      1039 tgaagaagaactctgtctgtcgtggaatcccggaacgttggtaactgtctcaggtagtg 1098
Oy      1094 tatgtaatccaagaacgacgacggtgtatcaactaacaatggaatctgcaagaca 1153
Db      1099 tatgtaatccaagaacgacgacggtgtatcaactaacaatggaatctgcaagaca 1158
Oy      1154 acaatgttaacgataatcgtttcatcggttccacgaatccaacaatcgttaaacctgg 1213
Db      1159 acaatgttaacgataatcgtttcatcggttccacgaatccaacaatcgttaaacctgg 1218
Oy      1214 ttgcttccaactgtgtacaatcgttcagatcgaacgcttccctcgcactctggtgtcctt 1273
Db      1219 ttgcttccaactgtgtacaatcgttcagatcgaacgcttccctcgcactctggtgtcctt 1278
Oy      1274 gggagttcatcccggtttagtgaacggttgggtggaacgtccgcgtgtaa 1320
Db      1279 gggagttcatcccggtttagtgaacggttgggtggaacgtccgcgtgtaa 1325

RESULT
7
AAV30576
ID AAV30576 standard; DNA; 1351 BP.
XX
AC AAV30576;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C gene in pHisBotA(syn).
XX
KW AntiToxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
KW botulinism; ds.
XX
OS Clostridium botulinum serotype A.
XX
FH Key Location/Qualifiers
FT CDS 1..1338
FT tag= a
FT
PN WO9808540-A1.
```

```
XX
PD 05-MAR-1998.
XX
PF 28-AUG-1997; 97WO-US15394.
XX
PR 28-AUG-1996; 96US-0704159.
XX
PA (OPHI-) OPHIDIAN PHARM INC.
XX
PI Thalley BS, Williams JA.
XX
DR WPI: 1998-230234/20.
DR P-PSDB: AAW68391.
XX
PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
XX
PS Example 29; Page 279-281; 428pp; English.
XX
CC This is the DNA sequence of the Clostridium botulinum serotype A
CC toxin C fragment gene contained in plasmid pHisBotA(syn). The
CC encoded toxin A polypeptide (see AAW68391) has a histidine-tagged
CC N-terminal extension. The vector was used to express native
CC (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC cells. The invention relates to recombinant proteins derived from
CC C. botulinum toxins. Methods are provided which allow for the
CC isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells.
CC The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
XX
SQ Sequence 1351 BP; 407 A; 348 C; 247 G; 349 T; 0 other:

Query Match 98.2%; Score 1302.2; DB 19; Length 1351;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 14 ttgcttctactctcactgaatatacataagaactccaataactcatctgaactcgc 73
Db 32 ttgtctgtactcttcaactgaatatacataagaactccaataactcatctgaactcgc 91
Oy 74 gctacgaatccaatcacctgatcgactgtctgcgtacagcttccaaatcaaatcggtt 133
Db 92 gctacgaatccaatcacctgatcgactgtctgcgtacagcttccaaatcaaatcggtt 151
Oy 134 ctaaatgaacttcgatccgatcgcagaagaatcagaatccagctgttccaatctgaaatc 193
Db 152 ctaaatgaacttcgatccgatcgcagaagaatcagaatccagctgttccaatctgaaatc 211
Oy 194 ccaaaatcgaaatctatccctgaagaatgctatcgatatcaactctatgtaagaactctc 253
Db 212 ccaaaatcgaaatctatccctgaagaatgctatcgatatcaactctatgtaagaactctc 271
Oy 254 ccaactctcttggatccgatccgaaatcactccaactccatctctctgaacaatgaat 313
Db 272 ccaactctcttggatccgatccgaaatcactccaactccatctctctgaacaatgaat 331
Oy 314 acacccatcatcaactgcatggaagaacaatctgttggaaagctatctctgaactaggtg 373
Db 332 acacccatcatcaactgcatggaagaacaatctgttggaaagctatctctgaactaggtg 391
Oy 374 aaatcttgtagctctgcaagacacacccaagaatccaacgctgtgttcaaatact 433
Db 392 aaatcttgtagctctgcaagacacacccaagaatccaacgctgtgttcaaatact 451
Oy 434 ctcaatgatcaacatctcgcatacatcaatcgcgtgactctgtttaccatcaccaaca 493
Db 452 ctcaatgatcaacatctcgcatacatcaatcgcgtgactctgtttaccatcaccaaca 511
```


OY	494	atcgtctcgtataaactccaataatcatcaaacaggccgctctgtatcgcaccagaacccgactc	553
Db	512	atcgtctcgtataaactccaataatcatcaaacaggccgctctgtatcgcaccagaacccgactc	571
OY	554	ccaatcgcggataaatccacagctcccaataacatcgaatgtcaaatctgagcggttgcgtg	613
Db	572	ccaatcgcggataaatccacagctcccaataacatcgaatgtcaaatctgagcggttgcgtg	631
OY	614	acaatcacccgtacatctcgtatcatcaaatctcaatctgtctgcacaaagaactcgtacagaa	673
Db	632	acaatcacccgtacatctcgtatcatcaaatctcaatctgtctgcacaaagaactcgtacagaa	691
OY	674	aagaatccaagaagacctctgtacgcacaaccagtcaccaatctgtatctcctcgtaaagactctgg	733
Db	692	aagaatccaagaagacctctgtacgcacaaccagtcaccaatctgtatctcctcgtaaagactctgg	751
OY	734	gtgcactccctgcgcgttaagacaacacgttactaatctgttaattctgtacatcccgacaacat	793
Db	752	gtgcactccctgcgcgttaagacaacacgttactaatctgttaattctgtacatcccgacaacat	811
OY	794	acgtttgcacgtcaacaagaatcgtgcatccgcggttaacatgtacactcgtgaagaagtcocgctgg	853
Db	812	acgtttgcacgtcaacaagaatcgtgcatccgcggttaacatgtacactcgtgaagaagtcocgctgg	871
OY	854	ctgttatgtactaccacaatactaccctgaactctccctcgtacaccgctgtgatccaatatcatca	913
Db	872	ctgttatgtactaccacaatactaccctgaactctccctcgtacaccgctgtgatccaatatcatca	931
OY	914	tcaagaataatccgcgtctcgtgttaacaagaacataatcgtttcgcacaacatgatcgtgtataca	973
Db	932	tcaagaataatccgcgtctcgtgttaacaagaacataatcgtttcgcacaacatgatcgtgtataca	991
OY	974	tcaatgtgtcgttaagaacaagaagaataccgcctcgtcgtacaaatgctctcctcagcgtcgtg	1033
Db	992	tcaatgtgtcgttaagaacaagaagaataccgcctcgtcgtacaaatgctctcctcagcgtcgtg	1051
OY	1034	tagaaaagactctgtctcgtctcgtgaaatccccggagcgtgtgttaatctcgtctcagtgatcg	1093
Db	1052	tagaaaagactctgtctcgtctcgtgaaatccccggagcgtgtgttaatctcgtctcagtgatcg	1111
OY	1094	taatgaataatccaagaagaagaccagggatatacctaacaatcgaataatgatctcgcacagaca	1153
Db	1112	taatgaataatccaagaagaagaccagggatatacctaacaatcgaataatgatctcgcacagaca	1171
OY	1154	acaatgtgaacgatatacgtttcatccggtttccacacagttcaacaatatcgcctaactcg	1213
Db	1172	acaatgtgaacgatatacgtttcatccggtttccacacagttcaacaatatcgcctaactcg	1231
OY	1214	ttgcttccaactcgtgtacaaatcgcatacgaatcgaacggttctcctctcgcacactcgtggttcctt	1273
Db	1232	ttgcttccaactcgtgtgtacaaatcgcatacgaatcgaacggttctcctctcgcacactcgtggttcctt	1291
OY	1274	gggaggttcatcccggttgcgtatgaacggttggggtgaaacgcttcgcctgttaa	1320
Db	1292	gggaggttcatcccggttgcgtatgaacggttggggtgaaacgcttcgcctgttaa	1338
RESULT 8			
AAT29246			
ID	AAT29246 standard; DNA; 1402 BP.		
XX			
AC	AAT29246;		
XX			
DT	07-JUL-1996 (first entry)		
DE	Type A neurotoxin C fragment-polymistidine tag gene fusion.		
XX			
KM	Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;		
XX	Clostridium botulinum; polynistidine; vector; PERHSA; phisbot; ds.		
OS	Synthetic.		
XX			
PH	Key	Location/Qualifiers	

FT	CDS	1..1317
FT		/*tag= a
FT		/product= phisBot fusion protein
PN	MO9612802-A1.	
XX		
XX	02-MAY-1996.	
PD		
XX	23-OCT-1995;	95WO-US13737.
XX		
PR	07-JUN-1995;	95US-0480604.
PR	24-OCT-1994;	94US-0329154.
PR	16-MAR-1995;	95US-0405496.
XX	14-APR-1995;	95US-0422711.
XX		
PA	(OPHI-) OPHIDIAN PHARM INC.	
XX		
PI	Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;	
PI	Williams JA;	
DR	WPI; 1996-230603/23.	
DR	P-PSDB; AAR95009.	
XX		
PT	Fusion proteins comprising non-toxin protein and part of toxin	
PT	useful to form anti-toxins against Clostridium botulinum type A, and	
PT	C. difficile type toxins, and to treat C. difficile intoxication,	
PT	partic. diarrhoea	
XX		
PS	Example 24; Page 340-342; 434pp; English.	
XX		
CC	A nucleotide sequence (AAR79246) present in vector pETHisa encodes	
CC	the phisBot fusion protein (AAR95009) comprising a polystidine	
CC	affinity tag and fragment C (see also AAR95008) of the Clostridium	
CC	botulinum type A neurotoxin. The phisBot protein was expressed	
CC	in Escherichia coli as a soluble protein and was purified by	
CC	metal chelate affinity chromatography to obtain a product free	
CC	of endotoxin contamination that may be useful as an immunogen	
CC	in vaccine compns.	
XX		
SO	Sequence 1402 BP; 420 A; 360 C; 260 G; 362 T; 0 other:	
	Query Match	98.2%; Score 1302.2; DB 17; Length 1402;
	Best Local Similarity	99.8%; Pred. No. 0;
	Matches 1304; Conservative	0; Mismatches 3; Indels 0; Gaps 0;
QY	14 tggctctaccttcaatgatacatcacaagaacatcatcaataccttccatctggaactgc	73
DB	83 tggctctaccttcaatgatacatcacaagaacatcatcaataccttccatctggaactgc	142
QY	74 ggtacgaatccatcaactcgatcgagcctgtctcgcgtacgcttccaaatacgaactcggtt	133
DB	143 ggtacgaatccatcaactcgatcgagcctgtctcgcgtacgcttccaaatacgaactcggtt	202
QY	134 ctaaaagttaactcgatccgcatcgacgaacgaatccagatccagctgttcaatctggaactc	193
DB	203 ctaaaagttaactcgatccgcatcgacgaacgaatccagatccagctgttcaatctggaactc	262
QY	194 ccaaaatcgaagttatctctgtaagaatgctatcgtatatacaactctatgtacgaaaactct	253
DB	263 ccaaaatcgaagttatctctgtaagaatgctatcgtatatacaactctatgtacgaaaactct	322
QY	254 ccaactccttctggatccgtaaccggaatatttaactcactctctcctggaacaatgat	313
DB	323 ccaactccttctggatccgtaaccggaatatttaactcactctctcctggaacaatgat	382
QY	314 acacacatcaactcgatcgagaacaaactctcggtttggaagatlatcttgaactacggtg	373
DB	383 acacacatcaactcgatcgagaacaaactctcggtttggaagatlatcttgaactacggtg	442
QY	374 aaatcaatctggaactctgcggagcaactcaggaatacaaacagcggtgtgatcaaatct	433
DB	443 aaatcaatctggaactctgcggagcaactcaggaatacaaacagcggtgtgatcaaatct	502

Db 383 acaccatcaatcgtcgtggaacaaatctgttggaagtactctctgtaactcgtg 442
 QY 374 aaatcatctgactctgcagagacactcaggaatacacaacgctgtgtatcaaatc 433
 Db 443 aaatcatctgactctgcagagacactcaggaatacacaacgctgtgtatcaaatc 502
 QY 434 ctcatgatacaatctctgatactacaacacgctgtatctctgtatcaacacacaa 493
 Db 503 ctcatgatacaatctctgatactacaacacgctgtatctctgtatcaacacacaa 562
 QY 494 atgctcgtataatcccaaaatctacatcaacgctgtcgtatcgacacagaaccgatc 553
 Db 563 atcgtcgtataatcccaaaatctacatcaacgctgtcgtatcgacacagaaccgatc 622
 QY 554 ccaatcgtgtataatcccaacgctcttaataacatcattcaaatctgacgctgtgtcgt 613
 Db 623 ccaatcgtgtataatcccaacgctcttaataacatcattcaaatctgacgctgtgtcgt 682
 QY 614 aaactcacgctacatctgatactacaactcaactctgtctgcacaaagaaactgaaagaa 673
 Db 683 aaactcacgctacatctgatactacaactcaactctgtctgcacaaagaaactgaaagaa 742
 QY 674 aagaatacaaaagacgtctgacacaaacagtcacattctgtatcctcgtaaagactctggt 733
 Db 743 aagaatacaaaagacgtctgacacaaacagtcacattctgtatcctcgtaaagactctggt 802
 QY 734 gtgactaccgtcagtaacgacaacacgctacatcgtcgtatcgtatcagacacgaacat 793
 Db 803 gtgactaccgtcagtaacgacaacacgctacatcgtcgtatcgtatcagacacgaacat 862
 QY 794 acgttgacgtcaacaatgtatgtaaccggttaccatgtacatcgtgaaagtcgcgctgtgt 853
 Db 863 acgttgacgtcaacaatgtatgtaaccggttaccatgtacatcgtgaaagtcgcgctgtgt 922
 QY 854 ctgttatgactaccacaatctcactgtaactctccctgtacacgctgtgcacaaatcaca 913
 Db 923 ctgttatgactaccacaatctcactgtaactctccctgtacacgctgtgcacaaatcaca 982
 QY 914 tcaagaataacgctcgtgtacaaagaaacatcgtcttcgcaacaaatcgtgtatca 973
 Db 983 tcaagaataacgctcgtgtacaaagaaacatcgtcttcgcaacaaatcgtgtatca 1042
 QY 974 tcaatgtgtatgtaagaacaagaataccgctcgtgtacaaatcgtctcagagctgtgt 1033
 Db 1043 tcaatgtgtatgtaagaacaagaataccgctcgtgtacaaatcgtctcagagctgtgt 1102
 QY 1034 tagaagaatctctgtcgtctgtgaatacccgacgctgtgtatcctgtctcagatgtgt 1093
 Db 1103 tagaagaatctctgtcgtctgtgaatacccgacgctgtgtatcctgtctcagatgtgt 1162
 QY 1094 taatgaataccaagaacagacaggtatcactaaacaaatgtaaaatgtaactgcagagca 1153
 Db 1163 taatgaataccaagaacagacaggtatcactaaacaaatgtaaaatgtaactgcagagca 1222
 QY 1154 acaatgttaacgataatcgtgttccatcgttccacacagttcaacaataatcgtcaaacg 1213
 Db 1223 acaatgttaacgataatcgtgttccatcgttccacacagttcaacaataatcgtcaaacg 1282
 QY 1214 ttgcttccaactcgttaacaatcgtcagatcgaaagctcctcctcgaactcgtgtgtctt 1273
 Db 1283 ttgcttccaactcgttaacaatcgtcagatcgaaagctcctcctcgaactcgtgtgtctt 1342
 QY 1274 gggaggtcattccggttgaatgagcgttggttgaaagtcgcgctgtaa 1320
 Db 1343 gggaggtcattccggttgaatgagcgttggttgaaagtcgcgctgtaa 1389

RESULT 10
 AAA54588
 ID AAA54588 standard; DNA: 1332 BP.
 XX
 AC AAA54588;
 XX

DT 11-APR-2001 (first entry)
 XX
 DE Sequence encoding botulinum toxin C fragment (serotype A).
 XX
 KW Botulinum; toxin; neurotoxin; heavy chain; recombinant expression;
 KW recombinant vector; antigen; immune response; vaccine; bacterium;
 KW infection; ds.
 OS
 OS Clostridium botulinum.
 XX
 PN W020067700-A2.
 PD 16-NOV-2000.
 PE 12-MAY-2000; 2000WO-US12890.
 XX
 PR 12-MAY-1999; 99US-0133865.
 PR 12-MAY-1999; 99US-0133866.
 PR 12-MAY-1999; 99US-0133867.
 PR 12-MAY-1999; 99US-0133868.
 PR 12-MAY-1999; 99US-0133869.
 PR 12-MAY-1999; 99US-0133873.
 PR 29-JUL-1999; 99US-0146192.
 XX
 PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
 PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
 XX
 DR WPI: 2001-016048/02.
 XX
 PT New nucleic acids encoding the carboxy- or amino-terminal portions of
 PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
 PT vaccine against botulinism
 XX
 PS
 PS Example 7; Page 37; 73pp; English.
 XX
 CC Botulinum neurotoxins are translated as a single 150 kDa polypeptide
 CC chain and then posttranslationally nicked, forming a dimer
 CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
 CC remain linked by a disulfide bond. Nucleic acids encoding the
 CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
 CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
 CC expression vectors and expressed in transformed cells to produce
 CC peptide antigens useful for eliciting an immune response to give
 CC protective immunity against botulinum neurotoxin, which causes
 CC botulinism. The nucleic acids are expressible in a recombinant
 CC organisms such as Escherichia coli or Pichia pastoris. The use
 CC of recombinant nucleic acids are advantageous since it eliminates
 CC the need to culture large quantities of hazardous toxin-producing
 CC bacterium. Production yield from the genetically engineered product
 CC is also high and cost of production is lower. The nucleic acids can
 CC be derived from Clostridium botulinum serotypes A-G.
 XX
 SQ Sequence 1332 BP; 401 A; 342 C; 244 G; 344 T; 1 other;

Query Match 94.1%; Score 1247.8; DB 22; Length 1332;
 Best Local Similarity 99.2%; Pred. NO. 0;
 Matches 1296; Conservative 1; Mismatches 4; Indels 6; Gaps 4;

QY 14 tggcctcactcactcaatcaatcaagaacatcaatcaatcctcgaactcgc 73
 Db 19 tgcgtctcactcactcaatcaatcaagaacatcaatcaatcctcgaactcgc 78
 QY 74 gctacgaatccaatcactgtagacctgtctcgtacgcttccaanaatcaacatcgtt 133
 Db 79 gctacgaatccaatcactgtagacctgtctcgtacgcttccaanaatcaacatcgtt 138
 QY 134 ctaaaagtaactcgtatcgatcgacgaagaatcaatcagctgtgttaactcgtgatctt 193
 Db 139 ctaaaagtaactcgtatcgatcgacgaagaatcaatcagctgtgttaactcgtgatctt 198

```

QY 194 ccaaaatcgagttatctctgaagaatgcatcgtatatacaactatgtagcgaactct 253
    |||
DB 199 ccaaaatcgagttatctctgaagaatgcatcgtatatacaactatgtagcgaactct 258
QY 254 ccaaccctctcgagtcgtatcccgaaatctcaactccatctctctgaacaatgat 313
    |||
DB 259 ccaaccctctcgagtcgtatcccgaaatctcaactccatctctctgaacaatgat 317
QY 314 acaccatcatcaactgtagcgaagaacaatctggttggaagatctctgaactagctg 373
    |||
DB 318 acaccatcatcaactgtagcgaagaacaatctggttggaagatctctgaactagctg 377
QY 374 aaatcatctcgagtcgtcgcgagacacacgaagaatacaacggtgtgttatccaact 433
    |||
DB 378 aaatcatctcgagtcgtcgcgagacacacgaagaatacaacggtgtgttatccaact 437
QY 434 ctccagatgtagcaacatctctgtagctatcaatcgcgtggtcttcgttaccatccaaca 493
    |||
DB 438 ctccagatgtagcaacatctctgtagctatcaatcgcgtggtcttcgttaccatccaaca 497
QY 494 atcgtctggaataactccaaaatctacatcaacggtcgtgtagcgcgaagaacgcatct 553
    |||
DB 498 atcgtctggaataactccaaaatctacatcaac-gccgctgtagcgcgaagaacgcatct 556
QY 554 ccaatctggtgtagacatccacgcgttcttaataacatcatgttcaaacctgtagctgt 613
    |||
DB 557 ccaatctggtgtagacatccacg-cttcaataacatcatgttcaaacctgtagctgt 615
QY 614 acactaccgctacatctctgtagcaaatcttcaatctgtctgcaagaagactggaagaa 673
    |||
DB 616 acactaccgctacatctctgtagcaaatcttcaatctgtctgcaagaagactggaagaa 675
QY 674 aagaaatcaaaagactgtagcgaacaacagtcgaatctgtagtacctgaagaactcttg 733
    |||
DB 676 aagaaatcaaaagactgtagcgaacaacagtcgaatctgtagtacctgaagaactcttg 735
QY 734 gtagtacctgtagcgaagaacacgtagtacctgtagtacctgtagtacctgtagtacct 793
    |||
DB 736 gtagtacctgtagcgaagaacacgtagtacctgtagtacctgtagtacctgtagtacct 795
QY 794 acgttgagctcaacaatgtagtlatccggttagcatgtagtacctgtagtacctgtagt 853
    |||
DB 796 acgttgagctcaacaatgtagtlatccggttagcatgtagtacctgtagtacctgtagt 855
QY 854 ctgttatgactccaacatctacactggaactctccctgtagtacctgtagtacctgtagt 913
    |||
DB 856 ctgttatgactccaacatctacactggaactctccctgtagtacctgtagtacctgtagt 915
QY 914 tcaagaataacgctgctgtagtcaagaagataatgctgtagcgaagaatgtagttagtaca 973
    |||
DB 916 tcaagaataacgctgctgtagtcaagaagataatgctgtagcgaagaatgtagttagtaca 972
QY 974 tcaatggtgtagttagaacaacaagataatcagctgtagcgaagaatgtagttagttagt 1033
    |||
DB 973 tcaatggtgtagttagaacaacaagataatcagctgtagcgaagaatgtagttagttagt 1032
QY 1034 tagaagaagatctgtagtctgtagtcaagaatcccgagactgtagttagttagttagt 1093
    |||
DB 1033 tagaagaagatctgtagtctgtagtcaagaatcccgagactgtagttagttagttagt 1092
QY 1094 taatgaataatccagaagacagggtagtacctaaacaaatgtagttagttagttagttagt 1153
    |||
DB 1093 taatgaataatccagaagacagggtagtacctaaacaaatgtagttagttagttagttagt 1152
QY 1154 acaatgtagaagataatcgtttagttagttagttagttagttagttagttagttagttagt 1213
    |||
DB 1153 acaatgtagaagataatcgtttagttagttagttagttagttagttagttagttagttagt 1212
QY 1214 ttgcttccaactgtagaactgtagcagtagcgaactgtagcagtagcagtagcagtagcag 1273
    |||
DB 1213 ttgcttccaactgtagaactgtagcagtagcgaactgtagcagtagcagtagcagtagcag 1272
QY 1274 gggagttcatcccggttagttagcaggttagtgggtgtagcagtagcagtagcagtagt 1320

```

```

DB 1273 gggagttcatcccggttagttagcaggttagtgggtgtagcagtagcagtagttaa 1319
    |||
RESULT 11
AAV30575
ID AAV30575 standard; DNA; 1546 BP.
XX
AC AAV30575;
XX
DT 07-DEC-1998 (first entry)
XX
DE Clostridium botulinum toxin A fragment C gene in phisBcA.
XX
KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;
XX botulism; ds.
XX
OS Clostridium botulinum serotype A.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 108..1496
XX FT /*tag= a
XX
XX MO9808540-A1.
XX
XX 05-MAR-1998.
XX
XX 28-AUG-1997; 97WO-US15394.
XX
XX 28-AUG-1996; 96US-0704159.
XX
XX (OPHI-) OPHIDIAN PHARM INC.
XX
XX Thalley BS, Williams JA;
XX
XX WPI: 1998-230234/20.
XX
XX P-PSDB: AAW68390.
XX
XX Host cell containing recombinant expression vector encoding
XX Clostridium botulinum type B or E toxin - useful to treat humans
XX and other animals at risk of intoxication with clostridial toxin
XX
XX Example 28; Page 277-278; 428pp; English.
XX
XX This is the DNA sequence of the Clostridium botulinum serotype A
XX toxin C fragment gene contained in plasmid pHisBcA. The encoded
XX toxin A polypeptide (see AAW68390) has a histidine-tagged
XX N-terminal extension. The vector was used to express native
XX (i.e. non-fusion) soluble C fragment in Escherichia coli host
XX cells. The invention relates to recombinant proteins derived from
XX C. botulinum toxins. Methods are provided which allow for the
XX isolation of soluble recombinant proteins free of significant
XX endotoxin contamination. Preferred hosts for production of
XX recombinant proteins are E. coli, insect cells and yeast cells.
XX The recombinant toxins are used as immunogens for the production
XX of vaccines and antitoxins that are useful in the treatment of
XX humans and animals at risk of intoxication with clostridial toxin.
XX
XX Sequence 1546 BP; 629 A; 163 C; 256 G; 498 T; 0 other;

```

Query Match 54.8%; Score 727.2; DB 19; Length 1546;
Best Local Similarity 72.2%; Pred. No. 1.2e-194;
Matches 945; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```

QY 19 tctacctaatcgataatcaagaacatcatcaactccatccatccgtagctgcctac 78
    |||
DB 195 tctacctaatcgataatcaagaacatcatcaactccatccatccgtagctgcctac 254
QY 79 gaatcaaatcactgtagcagctgtagcagctgtagcagctgtagcagctgtagcagctgtag 138
    |||
DB 255 gaatcaaatcactgtagcagctgtagcagctgtagcagctgtagcagctgtagcagctgtag 314

```

QY	139	gttaacttcgataccgatcgacaagaacatcgatccagctcgtgttcaatcctgaaatcttccaaa	198
Db	315	gtaaatcttgcatacgaatagaataaacaatccaattccaatttaatttgaagaagtagna	374
QY	199	atcgaaatttcctgaaagaattgcatacgtatacactatcgtataagaanaattctccacc	258
Db	375	atcgaggaattctttaaanaaatgctatttgtatataatagatagatgataaattttgaact	434
QY	259	tcctcttgcgataccgataatcccaatctccaactccatctctcctgaaacatgataacc	318
Db	435	agcctttggataagaattcctaagtattttaacagatataagctcctaataatgataataca	494
QY	319	atcataactgcatabganaaacaattcgtgttggaaagtaacctgaaactagcgttgaatc	378
Db	495	ataataatttgtatganaaataatttcagatgtagaagaatgaatacacttaattatgtgaaata	554
QY	379	atctgcgactctgcagagacactcagaatccaacacagcgtgtgtatccaataactccag	438
Db	555	atctgcgacttcaagagatactccgagaataaacaagaagtagtttttaaatcacgataca	614
QY	439	atgtgacaacctctgaactataccaatcgcgtcigatcttgcatacacaacaactcgt	498
Db	615	atgctataataatcacagttataataaacagatggattttctgataactatactaataataga	674
QY	499	ctgataactccaanaactctacatccaacggcgcgtcgtatgcacagaanaaccgataccat	558
Db	675	ttaaataacccttaaatattataataatggaattataatagatacaaaaaccaattccaat	734
QY	559	ctgggtacaatccaagcttctaataacatcatglttcaaatctgacagctgtgtcgtacat	618
Db	735	ttagtaataattcatcgtcgtagtaataataatgatttaaatgatgtgtgtagagataca	794
QY	619	caccgcacactcgtggtccaataatactccaatctgttcgcacaagaactgaacgaanaagaa	678
Db	795	catagataatcttgcgataaataattttaactctttgtgaataaggaatlaaatgnaaaaagaa	854
QY	679	atcaaaagacctgacgacaacacgctccaattcctgatactcctaagaactctcgggggtac	738
Db	855	atcaaaagatttatatgataataacatccaatctcagttattttaaagaacttttgggggtat	914
QY	739	taacctgcagtaacgaacaaaccgtactacatcaltgcgtgaatctgtacgactcgacaacaaatcgt	798
Db	915	tatttaacaatatgataaaccatactatgatttaaatttatatgatacgaataaataatgtc	974
QY	799	gacgtccaacatgtagttatccggcgtgtacacgtacactgaaagatgcgcggcgtgttcgtt	858
Db	975	gatttaataatgtaggtataggtataggtttatattgtatcttaaaaggcctagaaagtgcgtta	1034
QY	859	atgcataccaacactcaccggaactcttccctgcacggcgtgacccaataattccatacga	918
Db	1035	atgcataccaacatttcaattcaaatgttataggtaggggacaanaatttatataataa	1094
QY	919	aaatacgcgtctggtaaacaaggaacaatatcgttgcgaacaatgatacgtgtatataccat	978
Db	1095	aaatatcttctcggaataaagaataatattgttagaataatgatacgtatgatatattat	1154
QY	979	gttgtagttaagaacaagaataacgctctgcgtcacaatagctcttccaggctcgggtagaa	1038
Db	1155	gttagttaataaataaataaagaataataggttagctactaaatgcatacagccaggcgttagaa	1214
QY	1039	aaggtcttgcgtcctcgtgaaatcccgggacgtttgtgaaatcgttcacagtagtgttgaatg	1098
Db	1215	aaaaatactaatgctattagaataactctgtatgagaanaactcaagccaagtagtgaatcgt	1274
QY	1099	aaatccaagaagacagggatcactatacgaatgaaatgaaatcgtcagagacacat	1158
Db	1275	aagtcacaaaataatgatacgaagataaacaataaataatgaatatttaacaagataaatat	1334
QY	1159	gtataagatatcggtttcaatcggtttccacacagttcaacaatatcgtctaaactggtgtc	1218
Db	1335	gggataatgataagctttatagagatttccatcggttttaaataatattggtccaataatagca	1394

QY	1219	tccaactggtacatctgctaagatcgcaagcttctctcgcaccttggtgtctcttggag	1278
		11 11111 111 11 11111 11 11 111 111111111 11111	
Db	1395	agtaattgataataagacaataatagaataatctagtatagacctttggttctcataggaa	1454
QY	1279	ttcaatcccggttgatagacggttggtggatgcgcctcgtaagaatc	1326
		11 11 11 11111 11111 111 11 11111 11 11	
Db	1455	tttatctccgtatgatgatgtggtggagagaagaagcgaactgtaataatc	1502
RESULT	12		
AC	AAC64582		
XX	AAC64582 standard; cDNA; 4835 BP.		
XX	AAC64582;		
DT	15-FEB-2001 (first entry)		
XX			
DE	BONT/A neurotoxin prototoxin nucleotide sequence SEQ ID NO:10.		
XX			
KM	Human; procholecystokinin; CCK A receptor; CCK B receptor;		
XX	pancreatitis; antiinflammatory; ss.		
OS	Clostridium botulinum.		
PN	MO200061192-AZ.		
XX			
PD	19-OCT-2000.		
PF	06-APR-2000; 2000WO-US09142.		
XX			
PR	08-APR-1999; 99US-0288326.		
XX			
PA	(ALLR) ALLERGAN SALES INC.		
XX			
PI	Steward LE, Sachs G, Aoki KR;		
XX			
DR	WPI; 2000-679416/66.		
XX			
PT	New composition for treating acute pancreatitis, comprises a pancreatic		
PT	cell surface marker binding element, a translocation element that		
PT	transfers polypeptide across vesicular membrane, and a therapeutic		
PT	element .		
XX			
PS	Disclosure; Page 29-32; 50pp; English.		
XX			
CC	The present invention describes a composition (I) for treating acute		
CC	pancreatitis. (I) comprises a first element containing a binding element		
CC	that binds to a pancreatic cell surface marker, a second element		
CC	containing a translocation element that facilitates polypeptide transfer		
CC	across the vesicular membrane, and a third element containing a		
CC	therapeutic element that inhibits enzyme secretion in pancreatic cell		
CC	cytoplasm. Also described is a method for making a therapeutic		
CC	polypeptide having a binding element selective for cholecystokinin (CCK)		
CC	receptor by expressing within a host cell a recombinant chimeric		
CC	polypeptide comprising an extein containing a therapeutic element and a		
CC	translocational element, and an intein located to the carboxy terminal		
CC	of extein having a cysteine, serine or threonine at its amino terminus,		
CC	and contacting the extein with a synthetic peptide comprising a CCK		
CC	amino acid sequence containing an antideated phenylalanine at a natural		
CC	C-terminus, and a cysteine, serine or threonine at its N-terminus, and		
CC	a nucleophilic reagent able to cause cleavage of the intein to form a		
CC	peptide bond between the extein C-terminus and synthetic peptide		
CC	N-terminus through the formation of an activated ester or thio ester		
CC	intermediate. (I) has antiinflammatory activity and prevents accumulation		
CC	of pancreatic digestive enzymes, and prevents exocytic fusion of vesicles		
CC	containing secretory enzymes of pancreas. (I) is useful for treating		
CC	acute pancreatitis. The present sequence encodes the clostridium		
CC	botulinum BONT/A neurotoxin prototoxin which is given in the		
CC	exemplification of the present invention.		
XX			
XX	Sequence 4835 BP; 1934 A; 517 C; 756 G; 1628 T; 0 other;		

[illegible]

	DB	3967	aaataactgaagtcattagaataaacctcgtatgtaggaattcaagccaagtacgatg	4026
	OY	1099	aaatccaaaagaccagcggtatcacctaacaatgcgcaaatctgcaggacacaat	1158
	DB	4027	aagtcacaaaaatgatcataaggaataacaataatgcgcaaatgatcttacaagataaat	4086
	OY	1159	ggttaacgatatcggtttcatcggtttccaccaggttcaacaatatcgctaaactggtct	1218
	DB	4087	gggaatgatataagccttatagatttcacgtttaatatataagctaaactagtagca	4146
	OY	1219	tccaactggtacatcgttcagatcgcgaaccttcctccgcactctgggttgcttggag	1278
	DB	4147	agtaattggtataataatagacaaalagaaagcttcagtagagccttgggtctccaggaa	4206
	OY	1279	ttcacctccggtltgatagcgttgtgggtgaacgtccgcgtcgtgaagaatcc	1326
	DB	4207	tttatccctgtatgatgatgtagtgggagaagaagcgcaactgtaataatc	4254
	RESULT	13		
	AAZ87220			
	ID	AAZ87220	standard; DNA; 1317 BP.	
	XX	AAZ87220;		
	DE	08-MAY-2000	(first entry)	
	XX	DNA encoding native BONT serotype A (BONTA) C-terminal fragment (Hc).		
	XX	Botulinum neurotoxin; heavy chain; BONT; serotype A;		
	KM	C-terminal fragment; Hc; Venezuelan equine encephalitis virus replicon;		
	KW	VEE; botulism; vaccine; diagnosis; drug screening; ds.		
	XX	Clostridium botulinum.		
	XX			
	FH	Key	Location/Qualifiers	
	FT	CDS	1..1317	
	FT		/tag= "Native botulinum neurotoxin serotype A	
	FT		/product= (BONTA) heavy chain C-terminal fragment"	
	FT		/trans_except= (pos:1294..1314, aa:Gly)	
	PX	WO200002524-A2.		
	PN	20-JAN-2000.		
	PD			
	PF	09-JUL-1999;	99WO-US15570.	
	XX			
	PR	10-JUL-1998;	98US-0092416.	
	PR	12-MAY-1999;	99US-0133870.	
	XX	(USME-) US MEDICAL RES INST INFECTIOUS DISEASES.		
	XX			
	PI	Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;		
	DR	WPI: 2000-160827/14.		
	DR	P-PSDB; AAU77142.		
	XX			
	PT	Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum		
	PT	toxins serotypes A-G, is used for inducing an immune response against		
	XX	botulinum -		
	PS	Example 3; Page 52; 54pp; English.		
	CC	The invention relates to novel vaccines that induce a protective immune		
	CC	response against botulinum neurotoxin (BONT) serotypes A, B, C, D, E, F		
	CC	and G (BONTA-BONG). The vaccine of the invention is novel recombinant		
	CC	DNA construct comprising a vector, and at least one nucleic acid		
	CC	fragment comprising a C-terminal heavy chain fragment (Hc) from BONT		
	CC	serotypes A-G. In preferred embodiments of the invention, the vector is a		
	CC	Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this		
	CC	vector results in the production of large amounts of a protein encoded by		

CC a sequence cloned into the replicon. The constructs are used to produce
 CC vaccines against botulism. The proteins can also be used as diagnostic
 CC tools for the diagnosis of botulism. The transformed host cells can be
 CC used to analyse the effectiveness of drugs and agents which inhibit toxin
 CC effects. The vaccine currently used against botulism is dangerous
 CC and expensive to produce, and contains formalin, which is very painful
 CC for the recipient. Also, the vaccine is incomplete, in that only 5 of
 CC the 7 serotypes are represented in the formulation. The novel vaccine
 CC of overcomes these problems, as it is easily purified, and available in
 CC large quantities. It is also expressed in the lymph nodes for a better
 CC immune response. The present sequence represents DNA encoding native
 CC BoNTA heavy chain C-terminal fragment (Hc) used in an exemplification of
 CC the present invention.

CC
 XX
 SQ Sequence 1317 BP; 560 A; 108 C; 211 G; 438 T; 0 other;

Query Match 54.8%; Score 726; DB 21; Length 1317;
 Best Local Similarity 72.4%; Pred. No. 2,4e-194;
 Matches 942; Conservative 0; Mismatches 360; Indels 0; Gaps 0;

QY 19 tctaccttcactgatacatcaagaacatcctcaatcctccgaacccgctgcgtac 78
 DB 16 tctacattactgataatataaagataattatactctcattatgtaattagaatat 75
 QY 79 gaatccaataccgctgacgacctgtctgcgtacgcttccaataatccaatcggttctaa 138
 DB 76 gaaagtatcatcttaataagacttctcgtatgcatccaataataatattggtagtaaa 135
 QY 139 gtaacttcgactcgatcgacgaagaatcgaatccagctgttcaatctgaaatctccaaa 198
 DB 136 gtaaatcttgataccaatagataaataatccaattcattatattagaagtgtaaa 195
 QY 199 atcgaagtttcctggaagaatgctatcgtaacaaccttaagtgcgaataacttccacc 258
 DB 196 atggaagtaatttaaaaaaagctatgtaataatagtaatgtaagaattttagtact 255
 QY 259 tccctcggatccgtaaccgaataactccaatccatcctccgaacatgaatacacc 318
 DB 256 agccttggataaagaattcctcaagtaatttcaacgataaagctcaataatgaatatca 315
 QY 319 atcaacaatcgatgtaaaaaaactctggttgaagaatctcgaactacggtgaatc 378
 DB 316 ataataatgtatgtaaaaaaactcgaatggaagaatcacttaattatggtgaata 375
 QY 379 atcggactctgcgaacactcagaataatcaacagcgtgtgtatcttaataactctag 438
 DB 376 atcggactctgaagactcaggaataaataaagaagaagattttaataacagttcaa 435
 QY 439 atgtaacaatctctgatacatcaatcgtcgtatcgttaccatcaacaacatcgt 498
 DB 436 atgtaataatatacgaattataaacaagatgatttttgaactatcactcaataatga 495
 QY 499 ctgaataactccaataatctacatacagcgctgtgacgacgaacacgatatccaat 558
 DB 496 ttaataactcttaaatattataataatggaagattaaatgaataaaccatctcaat 555
 QY 559 ctgggtacatccacgcttcttaataacatcgttccaacgagcggtgtgcgtgacact 618
 DB 556 ttgaatataatctcgtcgaagtaataataatggttcaatattagatggtgtagagataca 615
 QY 619 caccgctacatctgatacaataactcaatcgttgcgaacgaacgaactgaaaaaaa 678
 DB 616 catagataatttggataaaaaatttcaatcttttggataaaggaattaaatgaaaaaaa 675
 QY 679 atcaagaactgtacgaacacagtcgaatctggtatccgtgaagactcttgggtgag 738
 DB 676 atcaaaatattatattgataatcaatcaatcaggtattttaaaagacttttgggtgag 735
 QY 739 taactggaagacgaacacgcttactataatggtgaatcgttcgaatccgaataatagct 798
 DB 736 tattacaataatgataaccacatactatattgtaaatcttaataatgatacgaataatattgc 795

QY 799 gacgtcaaatgtagatcatccggttaccatgtaacctgaagaagtcgcgctgttctgt 858
 DB 796 gatgtaaataatgtcgtatagatagaggttatatgtatcttcaaaagggccttaggtgcgta 855
 QY 859 atgataccaacatctacgtgaactctccctcgtacggtgtaacaaatcaatcaacga 918
 DB 856 atgataccaacatcttatttaattcaagttgttataggggacaaatattatataaa 915
 QY 919 aaatcgcgtctgtgaacaaaggaacatctgttcgaacaaatgacgtgtataatcaat 978
 DB 916 aaatcgtctgtgaacaaaggaacatctgttgaacaaatgacgtgtataatcaat 975
 QY 979 gttgagtgaagaacaaaggaacatccgtctgtaacaaatgacgtgtctcgaagcgtgtaga 1038
 DB 976 gtagtgaagtaaaaaataaagaatataatggttagctactaattgctcgaagcgtgtaga 1035
 QY 1039 aagatctgtcgtctcgtgaacatcccggaactgtgtaactcgtctcgaagttgttaag 1098
 DB 1036 aaatcgaagtgacatgaagaataacatcgtatgtagaagaatcgaagtagtaag 1095
 QY 1099 aaatcgaagaacgacgaaggtatcactcaacaaatgaaatgaatcgtcgaacacaaat 1158
 DB 1096 aagtcataaaatgacgaagaatgaacaaatgaatgaatgaatgaatgaatgaatgaat 1155
 QY 1159 gtaacgatacgtgttcaatcgttccacacggttcaacaaatcgtcgaagttgtct 1218
 DB 1156 ggaagtatagataggttctatagatgattcatcgaagtttaataatagctcaactagta 1215
 QY 1219 tccaactgttcaaatcgtcgaatcgaacggttccctcgaacatcgtgtgtctgtgag 1278
 DB 1216 agtaattgtataatagacaaatagaagaatcctagtagaacttgggtcgtcgaag 1275
 QY 1279 ttcatccggtgtgagcgtgttgggtgaagtcgcggtga 1320
 DB 1276 ttcatccggtgtgagcgtgttgggtgaagtcgcggtga 1317

RESULT 14
 AA87218 standard; DNA; 2532 BP.
 ID AA87218 standard; DNA; 2532 BP.
 AC AA87218;
 DT 08-MAY-2000 (first entry)
 XX
 DE DNA encoding native botulinum neurotoxin serotype A (BoNTA).
 XX
 KW Botulinum neurotoxin; heavy chain; BoNT; serotype A;
 KW Venezuelan equine encephalitis virus replicon;
 KW VEE; botulism; vaccine; diagnosis; drug screening; ds.
 XX
 OS Clostridium botulinum.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2532
 FT /tag= a
 FT /product= "Native botulinum neurotoxin serotype A
 FT (BoNTA) heavy chain"
 FT /transl_except= (pos:2509..2529, aa:GLY)
 FT
 PN WO200002524-A2.
 PD 20-JAN-2000.
 PD
 PD 09-JUL-1999; 99WO-US15570.
 PE
 PE 10-JUL-1998; 98US-0092416.
 PR 12-MAY-1999; 99US-0133870.
 PR
 PA (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
 PA Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
 PI
 XX

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 21:32:26 ; Search time 139.08 Seconds
(without alignments)
2341.890 Million cell updates/sec

Title: US-09-611-419a-5

Perfect score: 1326

Sequence: 1 gaattcgaacgatggcctc.....aacgtccgtctaagattc 1326

Scoring table: IDENTITY_NUC

Searched: Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCUTUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
	1	1302.2	98.2	1330	1	US-08-480-604A-22	Sequence 22, Appl
	2	1302.2	98.2	1330	2	US-08-405-496A-22	Sequence 22, Appl
	3	1302.2	98.2	1330	4	US-08-915-136-22	Sequence 22, Appl
	4	1302.2	98.2	1402	1	US-08-480-604A-25	Sequence 25, Appl
	5	1302.2	98.2	1402	2	US-08-405-496A-25	Sequence 25, Appl
	6	1302.2	98.2	1402	4	US-08-915-136-25	Sequence 25, Appl
	7	726	54.8	3891	1	US-08-480-604A-27	Sequence 27, Appl
	8	726	54.8	3891	2	US-08-405-496A-27	Sequence 27, Appl
	9	726	54.8	3891	4	US-08-915-136-27	Sequence 27, Appl
	10	166.4	12.5	1359	1	US-07-618-312A-3	Sequence 3, Appl
	11	166.4	12.5	1359	1	US-08-280-228-3	Sequence 3, Appl
	12	74.2	5.6	1359	1	US-07-618-312A-1	Sequence 1, Appl
	13	74.2	5.6	1359	1	US-08-110-786A-7	Sequence 7, Appl
	14	74.2	5.6	1359	1	US-08-280-228-1	Sequence 1, Appl
	15	74.2	5.6	1858	1	US-08-668-381A-6	Sequence 6, Appl
	16	60	4.5	5511	3	US-08-928-361B-2	Sequence 2, Appl
	17	60	4.5	7334	3	US-08-928-361B-1	Sequence 1, Appl
	18	57	4.3	5163	3	US-08-700-651-1	Sequence 1, Appl
	19	57	4.3	5163	3	US-08-928-361B-4	Sequence 4, Appl
	20	57	4.3	5318	3	US-08-700-651-2	Sequence 2, Appl
	21	57	4.3	5318	3	US-08-928-361B-3	Sequence 3, Appl
	22	40	3.0	1690	1	US-08-276-452A-24	Sequence 24, Appl
	23	40	3.0	1690	2	US-08-798-744-24	Sequence 24, Appl
	24	37.8	2.9	2277	1	US-08-676-967-2	Sequence 2, Appl
	25	37.8	2.9	2277	1	US-08-676-967-2	Sequence 2, Appl
	26	37.8	2.9	2277	2	US-09-098-487-2	Sequence 2, Appl
	27	37.2	2.8	18596	4	US-09-318-448-11	Sequence 11, Appl

28	36.8	2.8	1430	1	US-08-276-452A-25	Sequence 25, Appl
29	36.8	2.8	1430	2	US-08-798-744-25	Sequence 25, Appl
30	34.6	2.6	2094	3	US-08-714-918-87	Sequence 87, Appl
31	34.6	2.6	2094	4	US-09-265-315-87	Sequence 87, Appl
32	34.6	2.6	2094	4	US-09-265-315-87	Sequence 87, Appl
33	34.6	2.6	2094	4	US-09-265-315-87	Sequence 87, Appl
34	34.4	2.6	951	1	US-07-855-412B-2	Sequence 2, Appl
35	34.4	2.6	951	2	US-08-308-887A-2	Sequence 2, Appl
36	34.4	2.6	951	3	US-08-881-094-2	Sequence 2, Appl
37	34.2	2.6	198	5	PCT-US95-10668-1	Sequence 1, Appl
38	34.2	2.6	198	5	PCT-US95-10668-2	Sequence 1, Appl
39	34.2	2.6	198	5	PCT-US95-10668-3	Sequence 1, Appl
40	34.2	2.6	198	5	PCT-US95-10668-4	Sequence 1, Appl
41	34.2	2.6	3292	1	US-07-814-964-12	Sequence 12, Appl
42	34.2	2.6	3292	1	US-08-258-442-12	Sequence 12, Appl
43	34.2	2.6	3292	1	US-08-328-809-7	Sequence 7, Appl
44	34.2	2.6	3292	5	PCT-US92-11107-12	Sequence 12, Appl
45	34.2	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-22
; Sequence 22, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHAYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027


```

: FILING DATE: 31-OCT-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: INGOLIA, DIANE E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: OPHD-01308
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1330 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..1314
: US-08-405-496a-22

```

```

Query Match          98.2%; Score 1302.2; DB 2; Length 1330;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 14 tggccttacttactaatacatcaagaacatcatcaatccatccatcgaacctgc 73
Db 11 TGCGTCTACTCTTCACTAATACATCAAGAAACATATCATACCTCCATCTCGAACCTGC 70
QY 74 gctacgaatccaatcaatcgtatcgacctgtctcgctacgcttccaaaaatcaaatcgglt 133
Db 71 GCTAGATCCAAATCCATCCCTGATCGACCTGTCTGCTACGCTTCCAAATCAACATCGGTT 130
QY 134 cttaagttaacttgatcgcgatcgacaagaatcgaatccagctgtgtcaatctggaatctt 193
Db 131 CTAAAGTTAACTTCGATCCGATCGACAAAGAAATCAGATCCAGCTGTTCAATCTGGAATCTT 190
QY 194 ccaaatcgaagttaactcgtgaagaatgctatcgataacaactctatgttgaagaactctc 253
Db 191 CCAAAATGGAAGTTATCTCGAAGATGCTATGCTATACACCTCATGTGACGAACCTTCT 250
QY 254 ccaactcctctggaatccglatccgnaaatactcaactccatctctctgnaaataatg 313
Db 251 CCACCTCTTCTGATCCGTAATCCGGAATATCTTCAACTCCATCTCTGGAACATGAT 310
QY 314 acacatcataactgacatggaataaacttctgttgaagaatgatactcttgaactcgtgtg 373
Db 311 ACACATCATCAACTGCAATGGAACCAATCTGTTGGAAGATATCTCGAATCAAGGTG 370
QY 374 aaatcatctggaatcgtgacgaacatcaggaataacaacagcgtgtgtatcaaat 433
Db 371 AAATCATCTGACACTGTGAGGACACTCAGGAATCAAAACAGCGTGTGTGATTTCAATACT 430
QY 434 ctccagatgatacaactctctgactacatcaatcgtgtgacttctgttaccatcaaca 493
Db 431 CTCGATGATCAACATCTCTGACTCATCAATCCGCTGATCTTCTGTACATCACCACA 490
QY 494 atgtctgaataactccaaatctacatcaacagcgcgtctgatacgacaagaacccgalt 553
Db 491 ATGCTGTAATTAATCCAAATCTAATCAATCAACGCGCTGTGATCGACGAAGAACGATCT 550
QY 554 ccaatctggttaacatcaacgcttctaatacatcatgttcaaaatgtaagcgtgtgtg 613
Db 551 CCAATCTGGTAACTCAACGCTTCTAATATACATATGTTCAATGTGAGCGTGTGTG 610
QY 614 aaactacgcgtataatctgtgataaacttcaatctgttgcacaagaatcgaacga 673
Db 611 ACACCTACCGCTACTGATCTGATCAATCTCAATCTGTGACAAAGAACTGAACGAA 670
QY 674 aagaaataaagactgtgaacaaacccgltccaatctgtgtatccttgaagaactcttgg 733
Db 671 AAGAAATCAAGACCTGTGACGAACACGAGTCAATCTGTGATCTCTGAAAGACCTTGG 730

```

```

QY 734 gtaactacctgacgtacgacgaacacccgtactacatgctgtgactgtacgtccgaacaaat 793
Db 731 GTGACTACTGCGAGTACGCAACCCGTAATGCTGATGATCTGTACGATCCGAAACAAAT 790
QY 794 acgttgaagctcaacaatgtaggtatccgcggttacaatgtactgtgaagaagtcgggtgt 853
Db 791 ACGTTGACCTCAACAATGATGATGATCCGGGTTTACATGATGATGAAAGCTCCGCGGTG 850
QY 854 ctgttatgactcaacaactcactcgaactcttccctgacggtgtgtaacaaatcatca 913
Db 851 CTGTTATGACTACCAACATCTACTGACTCTTCCCTGATGAGGTGATGCAAAATTCATCA 910
QY 914 tcaagaataacgcgtctgttgaacaaagacaataatcgttccgaacaatgtatgtatata 973
Db 911 TCAAGAAATACCGCTGTGTGTAACAAGACAAATATGTTGCAACAATGATCGTGTATACA 970
QY 974 tcaatgtgttagttaaagaacaaatatacgcgtctgtgtacaaatgtcttcagcgtgtgt 1033
Db 971 TCAATGTTGTACTTAAAGAAACAAATACCGTGTGCTACCAATGCTTCTCAGGCTGTG 1030
QY 1034 tagaaagaatctgtctgtctggaatcccggaagctgtgtatctgtctcgaagtagtg 1093
Db 1031 TAGAAAGATCTTGTCTGCTGTGAAATCCCGAGCTGTGTATCTCTCAGGTACTTG 1090
QY 1094 taatgaatccaaagacagccaggtatcaactaacaataatgaataatcgtcagagaca 1153
Db 1091 TATATGAATCCAAAGAAACGACAGGATATCATACAAATGCAAAATGATCTGCAAGACA 1150
QY 1154 acaatgtaacgatalcgglttcaatcgtgttccaccagtticaaataatcgtctaaactgt 1213
Db 1151 ACAATGTTAATCGATATCGGTTTCAATCGGTTTCCACAGTTCAACATATCGCTAAACTGG 1210
QY 1214 ttgtcttccactggtgaacatcgtatgaatgaacgttccctcgtcgaactcgtgtgtctt 1273
Db 1211 TTGCTTCCAACTGTGTAACATCTCATGATGAAAGCTTCTCTGCACTGTGGGTGCTCTT 1270
QY 1274 gggagttcaccgggttgaacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1320
Db 1271 GGGAGTTCAATCCCGGTTGATGACGGTGTGGGTGGAACGTCGCGTGTAA 1317

RESULT 3
US-08-915-136-22
: Sequence 22, Application US/08915136
: Patent No. 6290960
: GENERAL INFORMATION:
: APPLICANT: KINK, JOHN A.
: APPLICANT: THALEY, BRUCE S.
: APPLICANT: PADHAY, NISHA V.
: APPLICANT: FIRCA, JOSEPH R.
: APPLICANT: STAFFORD, DOUGLAS C.
: TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
: PREVENTION OF C. DIFFICILE DISEASE
: NUMBER OF SEQUENCES: 32
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MEDLEN & CARROLL, LLP
: STREET: 220 MONTGOMERY STREET, SUITE 2200
: CITY: SAN FRANCISCO
: STATE: CALIFORNIA
: COUNTRY: UNITED STATES OF AMERICA
: ZIP: 94104
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/915,136
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/480,604
: FILING DATE:

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPMD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1330 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1314
US-08-915-136-22

Query Match 98.2% Score 1302.2; DB 4; Length 1330;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

14 tggcctctacatcctgatacatcaagaacatcaatcaatcctcctgaacctgc 73
11 tgcgtctacatcctgatacatcaagaacatcaatcctcctgaacctgc 70
74 gctgaacatcaatcctgatacatcaagaacatcaatcctcctgaacctgc 133
71 gctgaacatcaatcctgatacatcaagaacatcaatcctcctgaacctgc 130
134 ctaagatcaatcctgatacatcaagaacatcaatcctcctgaacctgc 193
131 ctaagatcaatcctgatacatcaagaacatcaatcctcctgaacctgc 190
194 ccaaatcgaagctacatcctgatacatcaagaacatcaatcctcctgaacctgc 253
191 ccaaatcgaagctacatcctgatacatcaagaacatcaatcctcctgaacctgc 250
254 ccaactcctcctgatacatcaagaacatcaatcctcctgaacctgc 313
251 ccaactcctcctgatacatcaagaacatcaatcctcctgaacctgc 310
314 acacatcaatcctgatacatcaagaacatcaatcctcctgaacctgc 373
311 acacatcaatcctgatacatcaagaacatcaatcctcctgaacctgc 370
374 aaatcctgatacatcctgatacatcaagaacatcaatcctcctgaacctgc 433
371 aaatcctgatacatcctgatacatcaagaacatcaatcctcctgaacctgc 430
434 ctcgaatcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 493
431 ctcgaatcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 490
494 atcgtcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 553

491 atcgtcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 550
554 ccaatcctgatacatcctgatacatcaagaacatcaatcctcctgaacctgc 613
551 ccaatcctgatacatcctgatacatcaagaacatcaatcctcctgaacctgc 610
614 acactcctgatacatcctgatacatcaagaacatcaatcctcctgaacctgc 673
611 acactcctgatacatcctgatacatcaagaacatcaatcctcctgaacctgc 670
674 aagaatcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 733
671 aagaatcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 730
734 gtaactcctgatacatcctgatacatcaagaacatcaatcctcctgaacctgc 793
731 gtaactcctgatacatcctgatacatcaagaacatcaatcctcctgaacctgc 790
794 acgtcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 853
791 acgtcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 850
854 cgtcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 913
851 cgtcgaatcctgatacatcaagaacatcaatcctcctgaacctgc 910
914 tcaagaatcctgatacatcaagaacatcaatcctcctgaacctgc 973
911 tcaagaatcctgatacatcaagaacatcaatcctcctgaacctgc 970
974 tcaatcctgatacatcaagaacatcaatcctcctgaacctgc 1033
971 tcaatcctgatacatcaagaacatcaatcctcctgaacctgc 1030
1034 tgaagaatcctgatacatcaagaacatcaatcctcctgaacctgc 1093
1031 tgaagaatcctgatacatcaagaacatcaatcctcctgaacctgc 1090
1094 taatgaatcctgatacatcaagaacatcaatcctcctgaacctgc 1153
1091 taatgaatcctgatacatcaagaacatcaatcctcctgaacctgc 1150
1154 acaatcctgatacatcaagaacatcaatcctcctgaacctgc 1213
1151 acaatcctgatacatcaagaacatcaatcctcctgaacctgc 1210
1214 tgcctcctgatacatcaagaacatcaatcctcctgaacctgc 1273
1211 tgcctcctgatacatcaagaacatcaatcctcctgaacctgc 1270
1274 ggaatcctgatacatcaagaacatcaatcctcctgaacctgc 1320
1271 ggaatcctgatacatcaagaacatcaatcctcctgaacctgc 1317

RESULT 4
US-08-480-604A-25
Sequence 25, Application US/08480604A
Patent No. 5736139
GENERAL INFORMATION:
APPLICANT: KINK, JOHN A.
APPLICANT: THALLEY, BRUCE S.
APPLICANT: PADHYE, NISHA V.
APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.
TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
PREVENTION OF C. DIFFICILE DISEASE
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA

```

COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,604A
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/422,711
FILING DATE: 14-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496
FILING DATE: 16-MAR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01763
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1386
US-08-480-604A-25

```

```

Query Match          98.2%; Score 1302.2; DB 1; Length 1402;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1304; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 314 acaccatcaactgcatgtaagaacaaatctggttggaagatctctgaaactgagtg 373
DB 383 ACACCATATCATCTGCAATGGAAGAAACATTTGGTTGGAAATATCTCGAACTACGGTG
QY 374 aaatcctggaactctgcaagcaactcaagaaatcaaacagcgttltgtatcaact 433
DB 443 AAATCATCTGCACTGCGAGACACTCAGGAATCAAAACAGCGTGTGATTCAAATACT
QY 434 ctcgaatgtaacatctctgactatcaatcgcgtggaatcttggttaaccacaca 493
DB 503 CTCAGATGATCAACATCTCTGACTCATATCAATTCGCTGGATCTTGATCCACCAACA
QY 494 atcgtctgataaccccaaatctacatcaacgagccgtctgacagcaaacatct 553
DB 563 ATCGTCTGATATCTCCAAATCTTACATTAAGGCCGTGTGATCGACCGAAACCGATCT
QY 554 ccaatctgggttaacatccaagctctcctaataacatcaatgtaaacatgga 613
DB 623 CCAATCTGGGTAAACATCCACGCTTCTAATPAATCATGTTCAAACTGACGGTGTG
QY 614 aacatccagcgtacatcttgatccaataacttcaatctgtctgcaaaaagacgaa 673
DB 683 ACACCTACCGCTACATCTGGATCAATTAATCTTCAATCTGTTGACAAAAGACGAA
QY 674 aagaaatcaaacgactgtacgaacaacccagtcgaattctgtatcctgaaagact 733
DB 743 AAGAAATCAAAACCTGTATCGAACMACAGTCCAATTTCTGTCTGAAAGACTTCTGG
QY 734 gtgaactcctgcaatgaagaacaaacgctactacatgctgaatctgtaaga 793
DB 803 GTGACTACCTGCAGTACGCAAAACCGTACTACATGCTGAATCTGTACGATCCGAACAAT
QY 794 agttgagctgaacaatgaggtatccggtgtaatgtaactgtaagaagtcggtg 853
DB 863 ACGTGAAGCTCAACATGAGGTATCCGGGTATCAATGTAACGTAAGAGTCCGCGGTT
QY 854 cgttatgactacaacatctaccgtgaactctccctgtaacgctgtgtaacaatcat 913
DB 923 CTGTATATGACTACCAACATCTACGTGAATCTTCCCTGATACGCTGTACCAATTCATCA
QY 914 tcaagaataacgctctgtaagaagacaatatcggttcgaacaatgactggtatata 973
DB 983 TCAGAAATATCCCGTCTGTGTAACAAGACAAATATGCTTGCAACATGATCGTATACA
QY 974 tcaatgttgtaagtaagaacaagaataaccgctggtgtaacaaatgcttcggtg 1033
DB 1043 TCAATGTGTAGTTAAGAACAAAGATATACGTTGGCTAACCAATGCTTCTCAGGCTGGTG
QY 1034 tagaaagaatctgtctgctcgtgaatccggaacgttgtaatcgtctcaggtagt 1093
DB 1103 TAGAAAAGATCTTGTCTGCTGTGGAATCCCGAGCTGTGTATCTGTCTCAGGTAGTTG
QY 1094 taatgaatccaagaacgaggtatcaactaacaatgcaaaatgaaatctgcaggaca 1153
DB 1163 TAATGAATCCAAAGAACGACAGGATACATCAACAAATGCAAAATGAAATCTGCGAGACA
QY 1154 acaatgtaagatagatcggttcatcaggtcttcacacagttcaacaatcgcgttaa 1213
DB 1223 ACAATGTATCAATATCGGTTTCAATCGGTTTCCACCAAGTTCAACATATCGTTAAACGTG
QY 1214 ttgcttcaaatgtaacaatcgtlcagatgaaacgttctcctcgtcgaactcgtggtt 1273
DB 1283 TTGCTTCAACATGATGATCAATGTCGATGCAATCGAATCGTTCCCTGCGACTCGGTTGCTCTT
QY 1274 gggagttcaatcccggttgatgaagcgttgggtggaacgtcgcgtgtaa 1320
DB 1343 GGGAGTTCAATCCCGGTTGATGACGGTGGGGAACGTCGCTGTAA 1389

```

```

RESULT 5
US-08-405-496A-25
; Sequence 25, Application US/08405496A
; Patent No. 5919665

```

```

1 GENERAL INFORMATION:
2 APPLICANT: WILLIAMS, JAMES A.
3 TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
4 TITLE OF INVENTION: NEUROTOXIN
5 NUMBER OF SEQUENCES: 30
6 CORRESPONDENCE ADDRESS:
7 ADDRESSEE: MEDLEN & CARROLL, LLP
8 STREET: 220 MONTGOMERY STREET, SUITE 2200
9 CITY: SAN FRANCISCO
10 STATE: CALIFORNIA
11 COUNTRY: USA
12 ZIP: 94104
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: Patent Release #1.0, Version #1.30
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/08/405,496A
21 FILING DATE: 16-MAR-1995
22 CLASSIFICATION: 424
23 PRIOR APPLICATION DATA:
24 APPLICATION NUMBER: US 08/329,154
25 FILING DATE: 25-OCT-1994
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 08/161,907
28 FILING DATE: 02-DEC-1993
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER: US 07/985,321
31 FILING DATE: 04-DEC-1992
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: US 07/429,791
34 FILING DATE: 31-OCT-1989
35 ATTORNEY/AGENT INFORMATION:
36 NAME: INGOLIA, DIANE E.
37 REGISTRATION NUMBER: 40,027
38 REFERENCE/DOCKET NUMBER: OPHD-01308
39 TELECOMMUNICATION INFORMATION:
40 TELEPHONE: (415) 705-8410
41 TELEFAX: (415) 397-8338
42 INFORMATION FOR SEQ ID NO: 25:
43 SEQUENCE CHARACTERISTICS:
44 LENGTH: 1402 base pairs
45 TYPE: nucleic acid
46 STRANDEDNESS: double
47 TOPOLOGY: linear
48 MOLECULE TYPE: DNA (genomic)
49 FEATURE:
50 NAME/KEY: CDS
51 LOCATION: 1..1386
52
53 US-08-405-496A-25

```

Query Match	98.2%	Score 1302.2	DB 2	Length 1402
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1304	Conservative	0	Mismatches 3	Indels 0
Oy	14	tgagctcactcactcaatacatcaacaatacatcaatcctccatcctgaacctgc	73	
Db	83	tgctctcacccttccactgatatcatcaagaatcatcatcattccatccctgaaacctgc	142	
Oy	74	gctacgaatccaatcaactcgtatcgaaactgctcgtcgaacgtctccaaatcaacaatcgtgt	133	
Db	143	gctacgaatcccatcaccctgacgcagctgctgcgtacccttccaaaatcaacatccggtt	202	
Oy	134	ctaaaggttaactcgatccgatccgacgaagaatccagatccagctgtgttcaatctggaattct	193	
Db	203	ctaaaggttaactcgtcatccgatccgacgaagaatcaaatccagctgtttcattctggaaattct	262	
Oy	194	ccaaaatcgaaattatcctgaaagatgcatctgtatatacaactctatgtacgaaaacttct	253	
Db	263	ccaaaatcgaaattatcctgaaagatgcatctgtattacaaactctatgtacgaaaacttct	322	

OY	254	ccacctcctcttgatctccgfatcccgaaatbaactbaacccatctctcttgaaaatgat	313
Db	323	ccacctctcttgatccgfatcccgaaatbaactbaacccatctctcttgaaacatgat	382
OY	314	aacacatcatcaactgcatactgaaacaactcttgcttgaaaglatctctgactacgctg	373
Db	383	acacatcatcaactgcatactgaaacaactcttgcttgaaaglatctctgactacgctg	442
OY	374	aaatcactctgactctcgaggaacaactcaggaaatcaaacgctgttgatccaatcatc	433
Db	443	aaatcattctggacgtctccagagacatcagaaaatcaaacggctgttgattccaattct	502
OY	434	ctcagatgatacaactctcgactcaactcaatcgcctggaatctctgataccatccaca	493
Db	503	ctcagatgatacaactctcgactcaactcaatcgcctggaatctctgataccatccaca	562
OY	494	atcgtctgataaactccaanaatctaatcaacgcccgtctgatacgaccagaaacgatct	553
Db	563	atcgtctgataaactccaanaatctaatcaacgcccgtctgatacgaccagaaacgatct	622
OY	554	ccaatctgggttaaaatcccaagcttctaatacaatcttgtaactgagacggttgctg	613
Db	623	ccaatctgggttaaaatcccaagcttctaatacaatcttgtaactgagacggttgctg	682
OY	614	aacatcacgcgtatcatcttgatcaaatbaactcaatctgtctgacaagaacatcgaa	673
Db	683	aacatcacgcgtatcatcttgatcaaatbaactcaatctgtctgacaagaacatcgaa	742
OY	674	aagaatacaaaagacctgtaagacaacccgctccaatctggtatccctgaaagactctgg	733
Db	743	aagaatacaaaagacctgtaagacaacccgctccaatctggtatccctgaaagactctgg	802
OY	734	gtgactacctgcgatacgaaaaacgctbaactaaatgctgtaactctgtaacatccgaacat	793
Db	803	gtgactacctgcgatacgaaaaacgctbaactaaatgctgtaactctgtaacatccgaacat	862
OY	794	acgttgacgtcaaaaagctgtagtaccgcggtgtacatgtaacctgaaaggtccgcgtgct	853
Db	863	acgttgacgtcaaaaagctgtagtaccgcggtgtacatgtaacctgaaaggtccgcgtgct	922
OY	854	ctgttatgataccaacaatacttaccitgaactctccctgtaacgctgtgtccaaatcatca	913
Db	923	ctgttatgataccaacaatacttaccitgaactctccctgtaacgctgtgtgtccaaatcatca	982
OY	914	tcagaanaatacgcgtctcgtgtaacaagaacatactgtctgcaacaatgatctgtatata	973
Db	983	tcagaanaatacgcgtctcgtgtaacaagaacatactgtgtgtgtgtgtgtgtgtgtgtgt	1042
OY	974	tcaatgtctgtagttaagaacaagaataacgcgtctggtctaccaaatgctctcaagctggtg	1033
Db	1043	tcaatgtctgtagttaagaacaagaataacgcgtctggtctaccaaatgctctcaagctggtg	1102
OY	1034	tagaaaagatctctgctgccttgaaatcccggaactgtgatactctgctcaagctgctg	1093
Db	1103	tagaaaagatctctgctgccttgaaatcccggaactgtgatactctgctcaagctgctg	1162
OY	1094	laatgaaatccaagaacgaccagggtatcaactaaatgcaaatgaaatctgcagaca	1153
Db	1163	laatgaaatccaagaacgaccagggtatcaactaaatgcaaatgaaatctgcagaca	1222
OY	1154	acaatgtgaaagatactggttcatctggttccacgaagtgaacaatctgctaaactg	1213
Db	1223	acaatgtgaaagatactggttcatctggttccacgaagtgaacaatctgctaaactg	1282
OY	1214	ttgcttccaactgttacaactcgtcaagatcgaaacggttccctctcgaaactctggtgtgctct	1273
Db	1283	ttgcttccaactgttacaactcgtcaagatcgaaacggttccctctcgaaactctggtgtgctct	1342
OY	1274	ggagagctcatcccggttgatgacggttggtggtgaaacgctgcgctgtaa	1320
Db	1343	ggagagctcatcccggttgatgacggttggtggtgaaacgctgcgctgtaa	1389


```

1      RESULT      6      US-08-915-136-25
2      ; Sequence 25, Application US/08915136
3      ; Patent No. 6290960
4      ; GENERAL INFORMATION:
5      ; APPLICANT: KINK, JOHN A.
6      ; APPLICANT: THALLEY, BRUCE S.
7      ; APPLICANT: PADHYE, NISHA V.
8      ; APPLICANT: FIRCA, JOSEPH R.
9      ; APPLICANT: STAFFORD, DOUGLAS C.
10     ; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
11     ; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
12     ; NUMBER OF SEQUENCES: 32
13     ; CORRESPONDENCE ADDRESS:
14     ; ADDRESSEE: MEDLEN & CARROLL, LLP
15     ; STREET: 220 MONTGOMERY STREET, SUITE 2200
16     ; CITY: SAN FRANCISCO
17     ; STATE: CALIFORNIA
18     ; COUNTRY: UNITED STATES OF AMERICA
19     ; ZIP: 94104
20     ; COMPUTER READABLE FORM:
21     ; MEDIUM TYPE: Floppy disk
22     ; COMPUTER: IBM PC compatible
23     ; OPERATING SYSTEM: PC-DOS/MS-DOS
24     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
25     ; CURRENT APPLICATION DATA:
26     ; APPLICATION NUMBER: US/08/915,136
27     ; FILING DATE:
28     ; CLASSIFICATION:
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER: 08/480,604
31     ; FILING DATE:
32     ; PRIOR APPLICATION DATA:
33     ; APPLICATION NUMBER: US 08/405,496
34     ; FILING DATE: 16-MAR-1995
35     ; PRIOR APPLICATION DATA:
36     ; APPLICATION NUMBER: US 08/329,154
37     ; FILING DATE: 25-OCT-1994
38     ; PRIOR APPLICATION DATA:
39     ; APPLICATION NUMBER: US 08/161,907
40     ; FILING DATE: 02-DEC-1993
41     ; PRIOR APPLICATION DATA:
42     ; APPLICATION NUMBER: US 07/985,321
43     ; FILING DATE: 04-DEC-1992
44     ; PRIOR APPLICATION DATA:
45     ; APPLICATION NUMBER: US 07/429,791
46     ; FILING DATE: 31-OCT-1989
47     ; ATTORNEY/AGENT INFORMATION:
48     ; NAME: INGOLIA, DIANE E.
49     ; REGISTRATION NUMBER: 40,027
50     ; REFERENCE/DOCKET NUMBER: OPD-01763
51     ; TELECOMMUNICATION INFORMATION:
52     ; TELEPHONE: (415) 705-8410
53     ; TELEFAX: (415) 397-8338
54     ; INFORMATION FOR SEQ ID NO: 25:
55     ; SEQUENCE CHARACTERISTICS:
56     ; LENGTH: 1402 base pairs
57     ; TYPE: nucleic acid
58     ; STRANDEDNESS: double
59     ; TOPOLOGY: linear
60     ; MOLECULE TYPE: DNA (genomic)
61     ; FEATURE:
62     ; NAME/KEY: CDS
63     ; LOCATION: 1..1386
64     ; US-08-915-136-25

```


Db 3490 AAATATGCTCTGCAAAATAAGATATATTTGATGAAATAATGATCGTATATATTAAT 3549
Qy 979 gttgaagtaagaacaagaataacgctctgctaccacatgctcccaagctggtgtagaa 1038
Db 3550 GTAGAGTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3609
Qy 1039 aagatctgtctgtctggaatacccgagcgttgtaactctgtcccaagtaactgta 1098
Db 3610 AAAATACAAAGATGATGAAATACCTGATGAGAAATCTAAGCAAGTACTACTAATG 3669
Qy 1099 aaatccaaagacgacgaggtatcctaacaatgcaaatgcaatctgcaggaacaacat 1158
Db 3670 AAGCAAAAATAATGATCAAGAAATGAAATGAAATGAAATGAAATGAAATGAAATG 3729
Qy 1159 ggtacagatcagctgttcacagctgttcacagctgttcacagctgttcacagctgt 1218
Db 3730 GGGATGATATAGCTTTATAGATTTGATGATTTGATGATTTGATGATTTGATGAT 3789
Qy 1219 tcccaactgtacacatcgtcagatcgaacgttcctcgcacactggtgctgtcgtg 1278
Db 3790 AGTATGTTGATTAATAGCAAAATAGAAAGATCTAGTACGCTTTGGTTGCTCATGG 3849
Qy 1279 ttcaatccggtgtgacggtgtggtggtggtggtggtggtggtggtggtggtggt 1320
Db 3850 TTTATTCCTGTAGATGATGATGGGAGAAAGCCACGTGTA 3891

RESULT 8

US-08-405-496a-27
Sequence 27, Application US/08405496a
Patent No. 5919665

GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.

TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
TITLE OF INVENTION: NEUROTOXIN

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP

STREET: 220 MONTGOMERY STREET, SUITE 2200

CITY: SAN FRANCISCO

STATE: CALIFORNIA

COUNTRY: USA

ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496a

FILING DATE: 16-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154

FILING DATE: 25-OCT-1994

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907

FILING DATE: 02-DEC-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321

FILING DATE: 04-DEC-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791

FILING DATE: 31-OCT-1989

ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.

REGISTRATION NUMBER: 40,027

REFERENCE/DOCKET NUMBER: OPND-01308

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 27:

SEQUENCE CHARACTERISTICS:
LENGTH: 3891 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3888
US-08-405-496a-27

Query Match 54.8%; Score 726; DB 2; Length 3891;
Best Local Similarity 72.4%; Pred. No. 1.3e-197;
Matches 942; Conservative 0; Mismatches 360; Indels 0; Gaps 0;

Qy 19 tctacctcaatgaaatcaagaacatcaatcaatcaatcaatcaatcaatcaatcaatcaat 78
Db 2590 TCTACATTTACTGAAATATTTAAGAAATTTAATTAATTTAATTAATTTAATTAAT 2649
Qy 79 gaatccaatcacctgtgacactgtctgcctgcctgcctgcctgcctgcctgcctgcct 138
Db 2650 GAAAGTAAATCAATTTAATGACTTATCTAGGTATGATCAAAATAATTAATTTGATG 2709
Qy 139 gtaactctgcac 198
Db 2710 GTAAATTTTGATGCAATATATTAATAATTAATAATTAATAATTAATAATTAATA 2769
Qy 199 atcgaagttatccctgaagaatcctacatcctacacacacacacacacacacacacac 258
Db 2770 ATTGAGGTAAATTTTAAAAATCTATTTGATTAATTAATTAATTAATTAATTAAT 2829
Qy 259 tccctgtgacatcgtatcccgaaatactccacatccatccctgtgacacacacacacac 318
Db 2830 ACCTTTTGATTAAGAAATCTTAAGTTTAAACAGTATTAAGTCTAATTAATTAATTA 2889
Qy 319 atcatcaactgacatgaagaacatccctgtgtgaaagatcctgtgacacacacacacac 378
Db 2890 ATAAATTAATTTGATGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2949
Qy 379 atctggaactctgcaggaac 438
Db 2950 ATCTGACCTTTACAGGATCTCAGGAAATAAACAAGAGTATTTTAAATACAGTCAA 3009
Qy 439 atgatacaatctctgac 498
Db 3010 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3069
Qy 499 ctgaataactccaatactcaacacacacacacacacacacacacacacacacacacac 558
Db 3070 TTAATTAATCTTAATAATTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3129
Qy 559 ctggtgaac 618
Db 3130 TTAGGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3189
Qy 619 caccgctacatctgac 678
Db 3190 CATAGATATATTTGATTAATTAATTTAAATTTTAAATTAATTAATTAATTAATTA 3249
Qy 679 atcaagacactgtac 738
Db 3250 ATCAAAAGATTTATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3309
Qy 739 tacctgcagtcagac 798
Db 3310 TATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3369
Qy 799 gacgtcaacaatggtggtatcccggttacatgactacacacacacacacacacacacac 858
Db 3370 GATGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 3429
Qy 859 atgactaccaacatctac 918

Db	3430	ATGACGTACAAACATTATTTAAATTCAGTTGTATAGGGGACAAATTTATTTATAAA	3489
Qy	919	aaatacgcctcgtglaacaagaacaalacgttcgcacaalatgcgtglatcaatcaat	978
Db	3490	AAATATCTCTCTGGAAATTAAGATATATTTGTAGAAATAAATGATCGTATATTTAAT	3549
Qy	979	gtgtagtagtaagaacaagaatccgtctcgttaccaatgcctcccaagtcggtgtagaa	1038
Db	3550	GTATATCTTAAAAATTAAGATATTTAGTTTAGCTACTAAATGATCAGGACAGGCGCTTAAA	3609
Qy	1039	aagaatctgcctcgtcgtgaatcccggaactgtgtgaatcgtctccagtagtgtaa	1098
Db	3610	AAATTACTAAGTGCATTAGAAATACCTGATGTAGCAAAATCTTAAGTCAAGTACTAATG	3669
Qy	1099	aaatccaagaacagcaggtgaltacactaaacaatgcaaatgaatcgtcagagcacacat	1158
Db	3670	AAGCAAAAAATGATCAGAGAAATACAAATTAATGCAAAATGAAATTTACAGATATATAT	3729
Qy	1159	ggttaacgatalcgtgttcatcgcgtttccacaggttcaacaalacgctaaactggtct	1218
Db	3730	GGGAATGATATAGGCTTTATATAGGATTTTCATGATTTAATATATATAGTAACTAGTACA	3789
Qy	1219	tccaactcgtgtacaatcgttcagatcgaaagttccctcgcacctcgtgtgtctctggag	1278
Db	3790	AGTATATGGTATATATAGCAAAATAGAAATATCTAGTAGGACTTTGGGTCTCATAGGGAA	3849
Qy	1279	ttaatcccggttgaatcaggtttgggtgtgaactcgcgtttaa	1320
Db	3850	TTTATTCCTTAGTGTAGTGTGAGTGGGAGAAAGGCCACTGTAA	3891

RESULT 9
US-08-915-136-27

; Patent NO. 6290960
; GENERAL INFORMATION:

APPLICANT: PADHYE, NISHA V
APPLICANT: THALLEY, BRUCE S
APPLICANT: KIRIN, JOHN A

APPLICANT: FIRCA, JOSEPH R.
APPLICANT: STAFFORD, DOUGLAS C.

```

; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
;
; NUMBER OF SEQUENCES: 32
;

```

ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 320 MONTGOMERY STREET SUITE 3200

;
;
CITY: SAN FRANCISCO
STATE: CALIFORNIA
;

ZIP: 94104
COMPUTER READABLE FORM:

```

; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: DOS 2.00 AND DOS 3.00

```

```

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

IN RE: FILING DATE: CLASSIFICATION:

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/405,496

;; FILING DATE: 16-MAR-1995
;; PRIOR APPLICATION DATA:
;; PUBLICATION NUMBER: 02 00 0000 1111

; FILING DATE: 25-OCT-1994
 ;
 ; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 05 06/101,50/
FILING DATE: 02-DEC-1993

```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/985,321
: FILING DATE: 04-DEC-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/429,791
: FILING DATE: 31-OCT-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: INGOLIA, DIANE E.
: REGISTRATION NUMBER: 40,027
: REFERENCE/DOCKET NUMBER: OPHD-01763
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 705-8410
: TELEFAX: (415) 397-8338
: INFORMATION FOR SEQ ID NO: 27:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3891 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 1..3888
: US-08-915-136-27

```

Query Match	54.8%;	Score 726;	DB 4;	Length 3891;
Best Local Similarity	72.4%;	Pred. No. 1.3e-197;		
Matches 942;	Conservative 0;	Mismatches 360;	Indels 0;	Gaps 0;

QY	19	lctacatccatctgataacatacaagaacatctcaataactccatccctgaaactcgcgcac	78
Db	2590	TCTACATTTACTGAAATATATTAAAGAAATATATTATACTCTATATTGAATTTAAAGATAT	2649
QY	79	gaatccaatcaaccgatacgcactgctcgcacgctctccaataatccaataatcggcttctaa	138
Db	2650	GAAAGTAACTCTTAAATATAGACTTATCTAGATGATGCATCAAAAATAAATATTTGGTAGTAA	2709
QY	139	gttaacttcgataccgatacgcagaagaatcagataccgactgctcaactctgaaactctccaa	198
Db	2710	GTAATTTTGTATCCATATGATATAAATCAATTCATATTATTATTATTTGAAGAAGTAGTAA	2769
QY	199	atcgaagatctccctgaagaatgctatcgatatacaactctatgtaacgaanaactctccacc	258
Db	2770	ATTGAGTAATTTTAAAAAATGCTATTTGTATATATACTAGTATGAAAAATTTTAGTACT	2829
QY	259	tcctcttgatcgcgataccgaaatcactccaccatcctctcgaaacatggaatacccc	318
Db	2830	AGCTTTTGGATGAGAAATTCCTTAGATTTTAAACGTAAAGCTTAATTAATTAAGATATATCA	2889
QY	319	atcatcaactcgcatacgaanaacaatctcgttctggaagaatctctcgaactcgcgtgaaatc	378
Db	2890	ATAATTAATTTGTATGGAATAATATTTACAGATGGAAGAATGATCACTTAATTTATGGCAATA	2949
QY	379	atctggaactctgcaaggacactccaaggaatccaacagcgltgttataatcaatctctag	438
Db	2950	ATCTGAGCTTTACAGGAACTCAGGAAATTAACAACAAAGATGATTTTAAATACAGTCAA	3009
QY	439	atgatacaactctcgcatacacaatcgcgcggatctcgtttacataccaacaacatcgt	498
Db	3010	ATGATTAATTAATATCGATTAATTAATTAACAAGATGGATTTTGTAACTATATCAATTAATATGA	3069
QY	499	ctgaataactccaanaatctacatacaacgcgcctcgtatcgcgacccaagaacccgactcccat	558
Db	3070	TTAATAATCTTAAATTTATATATAATGGAAGATTAATATAGATCAAAAACCAATTTCAANT	3129
QY	559	ctgggtaacacccaacgcttctaataacatcaltgttcaaatcgaacggtgtgcgtgacat	618
Db	3130	TTAGGTATATTCAATGCTAGTAATTAATATATGTTTAATTAATAGATGGCTTGACGATATCA	3189
QY	619	cacgcgtacacatctgatacaaatctactcaatctgltctgcgcaagaactcgaacgaanaagaa	678
Db	3190	CATGATATTTATTTGGATTAATATTTTATATCTTTTGGATGAAGAAATTAATGAAGAAACAA	3249

QY	679	atcaaacgactcgtacgacaaacccgacacgtcccaattcttgatccctccaaagactcttcggggtac	738
Db	3250	ATCAAAAGATTATATATGATATATCAATCAATTAATTTAGGTATTTTAAAGACTTTTGGGGTCAT	3309
QY	739	taccctgacgtacgacaaacccgtactacatgctgaaatctgtacgactcggacaaatacgtt	798
Db	3310	TATTTCAAAATATGATTAACACATCTATATGTTAAATTTATATGATCCAAATAAATATGTC	3369
QY	799	gaagctcaaacatgttagtataccgcggtttacatgatactccgaaaggtccgcggtggttcgtt	858
Db	3370	GATGTAAATATATGTAGGTATTAAGAGGTTTAAAGTTTAAAGGGCCATGAGGATAGGCTA	3429
QY	859	atgactccacaactctaccggaactcttcacctgacgtggtgacaaattcatcatcaag	918
Db	3430	ATGACGTCAAAACATTTATTTAAATTCOAATCTTGTTATAGGGGACAAATTTATTTATATAA	3489
QY	919	aaatacgcgtctggtatacaagaagcaatalcgttccgaacaatgatacgtgtatatacatcat	978
Db	3490	AAATATCTCTCTGGAAATTAAGATATATTTGTTAGAAATAAGATCGTGTATATATTAT	3549
QY	979	gttgtagtttagaacaagaataccggtctcgtgtaccaaatgctcttccaagctcgtgtgtgaa	1038
Db	3550	GTACTACTTAAAAATTAAGATATTAAGTTATAGCTACTAAATGATCAAGGACAGGCGTAA	3609
QY	1039	aagaatctgctcgtctcgtgaacatcccggaacgttgtgaaatcgtcttccaggtatgtgaa	1098
Db	3610	AAATATCTAGTGCATTAGAAATACCTGATGTGAGGAAATCTAAGTCAGTACTAGTATATG	3659
QY	1099	aaatccaagaagcaccagggtatcacttaacaaatgcnaaalyaatctgcaggaacaacat	1158
Db	3670	AAAGCAAAAAAATGATCAAGCAATAAACAAATAATGCAAAATGCAATTTTCAAGATATAT	3729
QY	1159	ggttaagaatactggtttcatcgtttccacacagttcaacaataatgcgtcaaaactggttct	1218
Db	3730	GGGAATATATTAGGCTTTTATAGATTTCATCAAGTTTAAATATATATGCTTAACCTAGTACA	3789
QY	1219	tccaactggtacaaatcgtccagatcgaagcttctctcgcactctggtgttcttctggag	1278
Db	3790	AGTAAATTTGTTATATAGCAAAATAGAAATCATCAGTAGAGACTTTGGTGTCTCAGGGA	3849
QY	1279	ttaatacccggttgatgaacggttggggtgaaagtcgcgtgtaa	1320
Db	3850	TTTATTTCTCTAGATGATGATGGATGGGAAGGACCATGTTAA	3891

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 8926832.0
 FILING DATE: 28-NOV-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9006097.1
 FILING DATE: 17-MAR-1990
 ATTORNEY/AGENT INFORMATION:
 NAME: Crawford M., Arthur R
 REGISTRATION NUMBER: 25,327
 REFERENCE/DOCKET NUMBER: 510-51
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 0101 703 8750400
 TELEFAX: 0101 703 5253468
 TELEX: 200797 NIXN UR
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1359 base pairs
 TYPE: NUCLEIC ACID
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Clostridium tetani
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..1356
 US-07-618-312A-3

	Query Match	12.5%	Score 166.4	DB 1	Length 1355	
	Best Local Similarity	52.3%	Pred. No. 5.2e-38			
	Matches 486	Conservative	0	Mismatches 411	Indels 33	Gaps 4
QY	43	aacatcatcaataactccatccctgaacactgctgctacgaaatcccaatccatccatctgacgctg	102			
Db	49	ATCCGTAAGAAAAGTCTACCACTTCTGAACTGTGGACATCAACAAAGAAATATTTCCTCCGCACAC	108			
OY	103	tctctgctacgctctccaaatcaacactcggttctaagttaacttcgactccg--atcgac	159			
Db	109	TCTGTTTCACACTCCTCTGTTATTCACATATTCAGATGCTCAATTGGTGGCGGACATCAAC	168			
OY	160	aagaatcgaatccagcgtgttcaatctggaatctccaaaatcgaaagttaactctgaaagt	219			
Db	169	GGCAAAGCATATCCACCTGGTTAAACAAGCATCTTGTGAAGTTATGCTGCACAAAGGCCATAG	228			
OY	220	gctatcgtataaactatctatgagaaactcttccactctctctgtagtcgtatcccg	279			
Db	229	GACATCGAATACAAAGACATGTTTCAACACTTCACCGTTAGCTTCTGGCTGGCGCTTCCG	288			
OY	280	aaatacttcaactctcatc-----tctctgaaacaatgaaatacaccaatcaactatgc	330			
Db	289	AAAGTTCTMGCTTCCACCTGGACCTGGACAGTAAAGGACACTAAAGAGTATCCATCATCTCT	348			
OY	331	atggaataacaat-----tctgtttgaaagtaactctctgtaactcaagtgaa	375			
Db	349	ATGAAGAAACACTTCCTCTGCATCGAGCTCTGTGGTTGGTCTGTTCCCTGGAAGGTTAAACAC	408			
OY	376	atcatctgagactctgacgaacactcaaggaataatcaacaacagcgtgtgtatcaataactct	435			
Db	409	CTGATCTGGACTCTGAAGAAGACTCGCGGGCGAAGTTTCGTCAATCACTTTCGGCGACCTG	468			
OY	436	cagatgatcaacaactctctgactacatcaaatcgctgtagctctcggttaaccaatcaacaat	495			
Db	469	CCGGCAAGATTCAAGCGCTACCTGGCTTAACAATAGGGTTTATCATCACTATCACTAAGCAT	528			
OY	496	cgtctgaaataactcaaaaatctcaatcaacagcgctctgactgacacgaacacgactctcc	555			
Db	529	CGTCTGTCTTCTGCTACACTGTACATCAAGCGCGTTTCTATATGGGCTCCGCTGAATATCACT	588			
OY	556	aatctggatcaaccatccagcttctataataacatcatglttcaaacctggaacggttgctcgtagc	615			

```

Db 589 GGTCTGGGCGCTTACCGTAGACACACATACATCTTAACTGACCGTTGCACACAC 648
QY 616 actcaccgctacatctggaataactcaatctctgacaaagaactgaacgaanaa 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AACACCAAGTACGTATCCGACACAGAGTTCCGTATCTTGCAAGACATCGAACCCGAA 708
QY 676 gaatacaagaactgtgacgaaccggtccaatctggtacctgaagaactctgggt 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 GAGATCGAAAAAGCTATACACACTACCTGTATCACTTCTGCTGCTGCTGCTGCTGCT 768
QY 736 gactaccgtgacgaacacccgctactacatgtgaaatctgtaacgacccaacacac 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AACCCGCTGGCTTACGACACACCGAATATTAACCTGATCCCGTACTGCTTAAAGAC 828
QY 796 gttgaacgtacaacatgtaagctacgcggttacatgtaacgtgaaggtccggtgtct 855
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 829 GTTACACTGAAAAA-----CATCACTGACTACATGACTGACCAACGGCGCTCTAC 882
QY 856 gttatgactacaaactctccgaactctccctgtaacgtgtgtacaaatcatc 915
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 ACTAACGTTAACTGAACATCTACTACCGACGTCTGTACAACGGCGCTGAATTCATCATC 942
QY 916 aagaatacgcgtcgtgtaacaagacaat 945
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 AACGCTACACTCCGAACACGAATCGAT 972

```

RESULT 11

US-08-280-228-3

Sequence 3, Application US/08280228

Patent No. 5571694

GENERAL INFORMATION:

APPLICANT: Makoff Dr, Andrew J

APPLICANT: Romanos Dr, Michael A

APPLICANT: Clare Dr, Jeffrey J

APPLICANT: Fairweather Dr, Neil F

TITLE OF INVENTION: VACCINES

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: NIXON & VANDERHUYE P.C.

STREET: 1100 No. 5571694th Glebe Road

CITY: Arlington,

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22201-4714

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/280,228

CLASSIFICATION: 435

FILING DATE: 25-JUL-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/618,312

CLASSIFICATION: 435

FILING DATE: 27-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 8926832.0

FILING DATE: 28-NOV-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9006097.1

FILING DATE: 17-MAR-1990

ATTORNEY/AGENT INFORMATION:

NAME: Wilson, Mary J.

REGISTRATION NUMBER: 32,955

REFERENCE/DOCKET NUMBER: 117-163

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NIXN UR

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-08-280-228-3

```

```

Query Match 12.5%; Score 166.4; DB 1; Length 1359;
Best Local Similarity 52.3%; Pred. No. 5.2e-38;
Matches 486; Conservative 0; Mismatches 411; Indels 33; Gaps 4;

QY 43 aacatcacaatcaccctcgaactcgtgacgaatcccaatccatcgaactcgt 102
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 49 ATCCTGAAAAAGTCTACCATTTGAACTTGACATCAACACGATATATCTCCGACATC 108
QY 103 tctcgctacgctcccaaatcaacatcgtgtccaagttaactctgacgcg---atgcac 159
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 109 TCTGTTTCAACTCTCTGTTATCATATTCAGATGCTCAATGTTGTCGGCGATCAAC 168
QY 160 aagaatcacatccagctgttcaatctgaaatctccaaatcgaaatctccgaagaat 219
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 169 GGCNAAGCTATCCACCTCGTTAAACAGAAATCTTGTGAATATCGTGCACAAAGCCATG 228
QY 220 gctatcgtatcaactctatgtaacgaactctccactcctctgtaacgtatccg 279
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 GACATCGAATACAAACGACATGTTCAACAACTGACGTTAGTCTGTCGGCGTTCCG 288
QY 280 aaatactcaactccatc-----tctctgaacaatgaataacacatcactcgc 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 AAAATTCTCTGCTTCCACCTGGAACAGTACGGCACTAACGAGTACCATCATCAGTCT 348
QY 331 atgaaatacaaat-----tctgtctggaagaatctctctgaactcaggtgaa 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 ATGAGAAACACTCCCTGCTCATCGGCTCTGTGGTCTTTTCCCTGAAGGGTAAACAC 408
QY 376 atcaatctgacatctgcaagcaactcaggaatcaaaacgcgtgtgtatcaaatcct 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 CTGATCTGACCTGGAAGACTCCGGCGGAGATTGCTGAGATCACTTCCGACACTG 468
QY 436 caaatgatacaatctctgactacatcaatcgcgtgatactgtgtacacaccacaat 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 CCGACAAAGTTCAACGCGTACCTGGCTTAAACAAATGGGTTTTCATCACTACTAACGAT 528
QY 496 cgtctgaataaactcaaaatctacaacgagcgctgtgacgaacgaacacgactcgc 555
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 529 GGTCTGCTTCTCTTAACCTGTTACATCAACGGCTTGTGATGGCTCCGCTGAATACAT 588
QY 556 aatctggttaacatccagctcttaataacatcatglttcaaatgacggtgtgac 615
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 589 GGTCTGGGCGCTATCCGCTGAGACAAACATCATCACTTAACTGACACCCGCAACAC 648
QY 616 actcaccgctacatctggaataactcaatctctgacaaagaactgaacgaanaa 675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 649 AACACCAAGTACGTATCCGACACAGAGTTCCGTATCTTGCAAGACATCGAACCCGAA 708
QY 676 gaatacaagaactgtgacgaaccggtccaatctggtacctgaagaactctgggt 735
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 709 GAGATCGAAAAAGCTATACACACTACCTGTATCACTTCTGCTGCTGCTGCTGCTGCT 768
QY 736 gactaccgtgacgaacacccgctactacatgtgaaatctgtaacgacccaacacac 795
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 769 AACCCGCTGGCTTACGACACACCGAATATTAACCTGATCCCGTACTGCTTAAAGAC 828
QY 796 gttgaacgtacaacatgtaagctacgcggttacatgtaacgtgaaggtccggtgtct 855

```



```

?
? TELEFAX: (703) 816-4100
? TELEX: 200797 NIKN UR
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1359 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA
? ORIGINAL SOURCE:
? ORGANISM: Clostridium tetani
?
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..1359
?
?
? US-08-110-786A-7

```

Query Match	5.6%	Score 74.2	DB 1	Length 1359
Best Local Similarity	47.3%	Pred. NO. 1.1e-11		
Matches 223	Conservative	0	Mismatches 248	Indels 0
				Gaps 0

QY	343	tcctgttgaaagatcctcgaactaagcgaaatctcggagctcggagcactcag	402
Db	376	TCCTGTTGGAGTGTATCACTTAAAGGTAATTAACCTTAATGACCTTTAAAGATTCCGG	435
QY	403	gaatccaacacgcgtgtgtatcattcaaatactcagaatgaatcaacatctgatacatc	462
Db	436	GGAGAGATTACCAATAAATTAACCTTTAGGAGATTACCTGAATTAATTAATGCTTATTAGCA	495
QY	463	aatgcgtcgagcttcgtttacatcaccacaacatcgtctgaaatcaatccaaatctacatc	522
Db	496	AATTAATGGGTTTTTATTAACATTACTTAATGAATGATTAATCTTCGTGTAATTTGTATATA	555
QY	523	aacggccgtctcgatccgacccaagaacgcgatcccaatctggtaacatccacgcgtcctaat	582
Db	556	AATGGAGACTTATTTGGGAAGTCGAGAATTAATTAAGTCTTTAGAGACTTATTAAGAGATTAAT	615
QY	583	aacatcatgltccaactcgaactcgagcgttgtctcgtaacatccacgcgtacatctgatacaat	642
Db	616	AATTAATCACTTAAACATTAAGATGATGTAATTAATTAATCAATCAATGCTTCTATTGATTAA	675
QY	643	ttcaatcgtctcgcgaagaagactgaagaagaagaatacaagaacctgttcgcgaacccg	702
Db	676	TTTTAGGATATTTTGCANAAGCAATTAAATCCAAAGATGTAATAAATTAATACACAGTTAT	735
QY	703	lccaactctgltatcctgaaagactcttgggttgatcactccgtgaatgaatcgacaaacggtat	762
Db	736	TTATCTATTAACCTTTTAAAGAGACTTCGCGGGAACCCCTTAGCATATGATACGAATAT	795
QY	763	tacatgctcgaaatctgtacgataccgaacaataatcgttgaagctcaacaatgta	813
Db	796	TATTTAATACCATGAGCTTCTAGTTCTTAATGAAGATGCTTCAATTTGAATAATTA	846

RESULT 14
US-08-280-228-1
Sequence 1, Application US/08280228
Patent No. 5571694
GENERAL INFORMATION:
APPLICANT: Makof Dr, Andrew J
APPLICANT: Romanos Dr, Michael A
APPLICANT: Clare Dr, Jeffrey J
APPLICANT: Fairweather Dr, Neil F
TITLE OF INVENTION: VACCINES
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: NIXON & VANDERHYE P.C.
STREET: 1100 NO. 5571694th Glebe Road
CITY: Arlington,
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22201-4714
COMPUTER READABLE FORM:

```

1 MEDIAN TYPE: Floppy disk
2 COMPUTER: IBM PC compatible
3 OPERATING SYSTEM: PC-DOS/MS-DOS
4 SOFTWARE: PatentIn Release #1.0, Version #1.25
5 CURRENT APPLICATION DATA:
6 APPLICATION NUMBER: US/08/280,228
7 FILING DATE: 25-JUL-1994
8 CLASSIFICATION: 435
9 PRIOR APPLICATION DATA:
10 APPLICATION NUMBER: US 07/618,312
11 FILING DATE: 27-NOV-1990
12 CLASSIFICATION: 435
13 PRIOR APPLICATION DATA:
14 APPLICATION NUMBER: GB 8926832.0
15 FILING DATE: 28-NOV-1989
16 PRIOR APPLICATION DATA:
17 APPLICATION NUMBER: GB 9006097.1
18 FILING DATE: 17-MAR-1990
19 ATTORNEY/AGENT INFORMATION:
20 NAME: Wilson, Mary J.
21 REGISTRATION NUMBER: 32,955
22 REFERENCE/DOCKET NUMBER: 117-163
23 TELECOMMUNICATION INFORMATION:
24 TELEPHONE: (703) 816-4000
25 TELEFAX: (703) 816-4100
26 TELEX: 200797 NIXN UR
27 INFORMATION FOR SEQ ID NO: 1:
28 SEQUENCE CHARACTERISTICS:
29 LENGTH: 1359 base pairs
30 TYPE: nucleic acid
31 STRANDEDNESS: double
32 TOPOLOGY: linear
33 MOLECULE TYPE: DNA (genomic)
34 HYPOTHETICAL: NO
35 ORIGINAL SOURCE:
36 ORGANISM: Clostridium tetani
37 FEATURE:
38 NAME/KEY: CDS
39 LOCATION: 1..1356
40 US-08-280-228-1
41
42 Query Match 5.6%; Score 74.2; DB 1; Length 1359;
43 Best Local Similarity 47.3%; Pred. No. 1.1e-11;
44 Matches 223; Conservative 0; Mismatches 248; Indels 0; Gaps 0;
45
46 343 tctgtgtggaagatctctctgaactacggtgaacatcatctggaactctgaagacactcaag 402
47 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
48 Db 376 tctgtgttgagagcgtatctactttaaagctaatatcttaattatgactctttaaagattcccg 435
49 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 Qy 403 gaatcaacaacgctgtctgtatctcaaatctctcaatctcagatgatcaacatctcgtactac 462
51 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
52 Db 436 GGAGAAAGTTAGACAATTAACCTTTAGGGATTTACTCGATTAATTTAAATGCTATTATGACA 495
53 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
54 Qy 463 aatgcctggaactctcgtttacatcccaacaatcgtctgaataactccaataactcaac 522
55 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56 Db 496 AATAAATATGGTTTATTAACCTATTACTATGATGATTAATTAATTAATTAATTAATTAATTA 555
57 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
58 Qy 523 aacgcgccttgatcgacacgaacacgcgatctccaatctcgaatctcgaatccaacgctctaat 582
59 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
60 Db 556 AATGGAGTACTTATATGGGAAGTGCAGAAATTAATCTGCTTTAGGAGCTATTAGAGAGATTAAT 615
61 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
62 Qy 583 aacatcatgttcaaacactgacggtctgctgacactcaacgcgtacatctgataccaataac 642
63 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
64 Db 616 AATAATTAACATTAAACATGATGATGATGATTAATTAATTAATTAATTAATTAATTAATTA 675
65 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 Qy 643 ttcactctgttcgacaaagaactgaaacgaaaaaaagaatcaaaagacctgtacgaacaacga 702
67 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
68 Db 676 TTTAGATATTATTTGCAAAAGCATTAATTAATCCAAAAGAGATTGAAAATTAATACACAAGTTAT 735
69 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
70 Qy 703 tccaattctgtatctctgaagaagactctcggggtgactaccctgcgaatgaacaacacgctac 762
71 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
72 Db 736 TTATCTAATTAACCTTTTAAAGACCTTTGGGAAACCCCTTAACGATTAATGATTAACAAATAT 795
73 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 16, 2002, 19:28:15 : Search time 4365.92 Seconds
(without alignments)
4099.242 Million cell updates/sec

Title: US-09-611-419A-5

Perfect score: 1326

Sequence: 1 gaattcggaacgagtcgctc.....aacgtcgcgtctaagaattc 1326

Scoring table:

IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gp_est1:*
10: gp_est2:*
11: gp_hic:*
12: gp_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.4	4.4	608	10	BI887904
2	56.2	4.2	1101	12	CNS00100
3	53.6	4.0	537	10	BM277996
4	53.6	4.0	551	10	BM278689
5	53.2	4.0	394	9	AU060224
6	52.6	4.0	447	10	BI594803
7	52	3.9	597	10	BM278181
8	51.8	3.9	500	10	BM278805
9	50.6	3.8	511	10	BM278805
10	50.4	3.8	569	10	BM278657
11	50	3.8	641	10	BM181884
12	49.6	3.7	645	9	AI389106
13	49.2	3.7	827	12	CNS02156
14	49	3.7	481	10	BF050030
15	48.8	3.7	628	9	AU060230
16	48.2	3.6	546	10	C22974
17	48.2	3.6	1101	12	CNS0100X

18	48	3.6	546	10	BF169335
19	48	3.6	556	10	BM277853
20	47.4	3.6	377	10	BI594814
21	47.4	3.6	450	9	AU060996
22	47	3.5	458	10	BI783045
23	47	3.5	518	10	BI782765
24	46.8	3.5	1175	10	BI872945
25	46.4	3.5	469	10	BF050033
26	45.4	3.4	386	10	BF050064
27	45.4	3.4	499	10	BM278692
28	45.4	3.4	536	10	BF050073
29	45.4	3.4	542	10	BM278786
30	45.4	3.4	554	10	C23753
31	45.2	3.4	501	10	BM277916
32	45.2	3.4	525	12	BM184460
33	45.2	3.4	525	12	CNS07PEA
34	45.2	3.4	1101	12	CNS000D1
35	45	3.4	564	10	BE606733
36	44.8	3.4	534	10	BM278087
37	44.6	3.4	1101	12	CNS016JY
38	44.4	3.3	681	12	BM278558
39	44.4	3.3	681	12	CNS02EOD
40	44.2	3.3	325	12	CNS03PK8
41	44.2	3.3	429	9	AU052930
42	44.2	3.3	500	12	BF7199
43	44.2	3.3	519	10	BM278164
44	43.8	3.3	535	10	BI594948
45	43.6	3.3	443	10	BM039802

ALIGNMENTS

RESULT 1
BI887904
LOCUS
DEFINITION
Zf637-1-002159 zebrafish shield stage whole embryo cDNA library
MPMGp637_Danio rerio cDNA clone MPMGp637_10F4:MPMGp637F0410 5',
mRNA sequence.
ACCESSION
BI887904
VERSION
BI887904.1
KEYWORDS
zebrafish.
SOURCE
Danio rerio
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 608)
Clark, M., Aanstad, P., Hennig, S., Johnson, S.L. and Lehrach, H.
EST sequencing of a zebrafish shield stage cDNA library normalised
by oligonucleotide fingerprinting
JOURNAL
Unpublished (2001)
COMMENT
Contact: Hennig S
Laboratory 123, dept. Lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inhest. 63-73, D-14195 Berlin, Germany
Tel: +49 30 8413 1612
Fax: +49 30 8413 1380
Email: hennig@molgen.mpg.de
5' EST sequencing of clones from a zebrafish shield stage library,
normalised from 55,000 starting clones by oligonucleotide
fingerprinting
High quality sequence stop: 608.
FEATURES
source
1. 608
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="MPMGp637_10F4:MPMGp637F0410"
/clone_lib="zebrafish shield stage whole embryo cDNA
library MPMGp637"
/tissue_type="whole embryo"
/dev_stage="shield stage, 6 hrs post-fertilisation"
/lab_host="E.coli, XLI blue MRF"

/note="Vector: pSport1; Site_1: NotI; Site_2: SalI;
 oligo-dr-NotI primed, SalI adaptors, directionally cloned,
 library normalised by oligonucleotide fingerprinting"
 BASE COUNT 207 a 218 c 41 g 140 t 2 others
 ORIGIN

Query Match 4.4%; Score 58.4; DB 10; Length 608;
 Best Local Similarity 49.1%; Pred. No. 0.00035;
 Matches 181; Conservative 0; Mismatches 187; Indels 1; Gaps 1;

QY 442 ataaacatctctgactacatcaa-tgcgtgactctcgttaacacacacatgctc 500
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 5 ATCAACTTCTCCAGTACATCAACAACTTCTCCAGTACATCACTTCTCCAGTACAA 64
 QY 501 gaataatccaataatcacatcaagcgctgcatcgacagcaaacgcttccatct 560
 || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 65 CAACAACTTCTCCAGTACATCACTTCTCCAGTACATCACTTCTCCAGTACAA 124
 QY 561 gggtaacatcacgctcttaataacatcatgcttaacatgagctgtgctgacatca 620
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 125 CATCACTTCTCCAGTACATCACTTCTCCAGTACATCACTTCTCCAGTACAA 184
 QY 621 ccgtacatctggtatcaaatctcaatctgctcgacaaagaactgacgaaaaagaat 680
 || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 185 CTCTCCAGTACATCACTTCTCCAGTACATCACTTCTCCAGTACAA 244
 QY 681 caaagacctgtacgaacacagtcgaatctgtatcctgaagactctgggggtgacta 740
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 245 CAACAACTTCTCCAGTACATCACTTCTCCAGTACATCACTTCTCCAGTACAA 304
 QY 741 cctgcagtagcagaacacgctactacatgctgtaactgtgacatcgacaaacatgctga 800
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 305 CATCACTTCTCCAGTACATCAACAACTTCTCCAGTACATCACTTCTCCAGTACAA 364
 QY 801 cgtcaacaa 809
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 365 CTACAAACA 373

RESULT 2
 LOCUS CDS00100 1101 bp DNA linear GSS 03-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
 BACR32D23 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL068607
 VERSION AL068607.1 GI:4958689
 KEYWORDS GSS.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutyo Osoegawa and
 Aaron Kammoser in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or

filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 location/Qualifiers

FEATURES
 source
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone_1lb="RPCI-98"
 /clone="BACR32D23"
 /note="end : TET3"
 BASE COUNT 155 a 166 c 7 g 284 t 489 others
 ORIGIN

Query Match 4.2%; Score 56.2; DB 12; Length 1101;
 Best Local Similarity 17.8%; Pred. No. 0.0016;
 Matches 101; Conservative 220; Mismatches 245; Indels 3; Gaps 1;

QY 17 cctctacccttaacgaatcacatcaagaacatcatcaatccatccatccatgacatcgct 76
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 496 MCTCYVACACATCCMIMACSMIMIMACSMIMIMACSMIMIMACSMIMIMACSMIMIM 555
 QY 77 agcaatccaatccatgactgactgctgctgactgactgactgactgactgactgactgact 136
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 556 AMCMITATCMIM 615
 QY 137 aagtaacttgatccatgacgaagaatgacatccagctgctgaatctggaatcttca 196
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 616 MCMITMTMATMIM 675
 QY 197 aaatcgaagtatccatgaagaatgactgactgactgactgactgactgactgactgact 256
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 676 MIM 732
 QY 257 cctccatctgactgactgactgactgactgactgactgactgactgactgactgactgact 316
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 733 CCCSCAMIM 792
 QY 317 ccatcatcaatgactgactgactgactgactgactgactgactgactgactgactgactgact 376
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 793 YMNIM 852
 QY 377 tcatctgactgactgactgactgactgactgactgactgactgactgactgactgact 436
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 853 TCHNTTTCMIM 912
 QY 437 agatgactcaatctgactgactgactgactgactgactgactgactgactgactgactgact 496
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 913 MIM 972
 QY 497 gctgtaataatccaaatcatcatcaagcgctgactgactgactgactgactgactgact 556
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 973 HTTTTMTTTCMIM 1032
 QY 557 atctgggtaacatccatgactgactgactgactgactgactgactgactgactgactgact 585
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 1033 MAM 1061

RESULT 3
 LOCUS BM277996 537 bp mRNA linear EST 20-DEC-2001
 DEFINITION As_192_51F02-SKPL Ascaris suum adult male testis germinal zone from
 Alan Scott Ascaris suum CDNA clone As_192_51F02 5', mRNA sequence.
 ACCESSION BM277996
 VERSION BM277996.1 GI:117971254
 KEYWORDS EST.
 SOURCE pig roundworm.
 ORGANISM Ascaris suum
 Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascaridoidea
 : Ascarididae; Ascaris.
 1 (bases 1 to 537)
 Blaxter M.L., Parkinson J., Whittton C., Daub J., Gulliano D., Hall
 N., Quayle M. and Barrell B.

TITLE	Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL	Unpublished (2000)
COMMENT	Contact: Blaxter ML

```

Db      342 ACAGTACTTCAGCACTTCAACAGCAACACTTCA 377
RESULT  5
LOCUS   AU060224
DEFINITION AU060224 Dictyostellium discoideum SL (H.Urushihara) Dictyostellium
ACCESSION AU060224
VERSION   AU060224.1 GI:4881328
KEYWORDS EST.
SOURCE   Dictyostellium discoideum.
ORGANISM Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
REFERENCE 1 (bases 1 to 394)
AUTHORS  Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
          Yoshino,R., Mitsu,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
          Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
          Unpublished (1998)
          Unpublished (1998)
          Contact: Hideko Urushihara
          Institute of Biological Sciences
          University of Tsukuba
          3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
          Email: d402huesakura.cc.tsukuba.ac.jp
          PROJECT = Dictyostellium discoideum cDNA project in Japan.
FEATURES
  source
    1..394
    /organism="Dictyostellium discoideum"
    /strain="AX4"
    /db_xref="taxon:44689"
    /clone_lib="SLA610"
    /dev_stage="slug"
BASE COUNT  223 a      134 c      8 g      28 t      1 others
ORIGIN
Query Match      4.0%; Score 53.2; DB 9; Length 394;
Best Local Similarity 49.3%; Pred. No. 0.0062;
Matches 139; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```

```

Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
; Ascarididae; Ascaris.
REFERENCE 1 (bases 1 to 447)
AUTHORS  Blaxter,M.L., Parkinson,J., Whittom,C., Daub,J., Gulliano,D., Hall
          N., Quayle,M. and Barrell,B.
          Edinburgh University/Sanger Centre Nematode EST Project
          Unpublished (2000)
          Contact: Blaxter ML
          Institute of Cell, Animal and Population Biology
          University of Edinburgh
          Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
          3JT, UK.
          Tel: +44 131 650 6760
          Fax: +44 131 670 5450
          Email: mark.blaxter@ed.ac.uk
          The library was prepared by Michelle Lizotte-Waniewski for Alan
          Scott, Johns Hopkins University Medical School, Baltimore.
          Sequencing was performed by the Pathogen Sequencing Unit, Sanger
          Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
          sequence contained a Polya tail (trimmed)
          PCR Primers
          FORWARD: T3
          REVERSE: T3
          BACKWARD: T7PL
          Plate: 44 row: C column: 11
          Seq primer: SKPL
          High quality sequence stop: 447.
          Location/Qualifiers
            1..447
            /organism="Ascaris suum"
            /db_xref="taxon:6253"
            /clone_lib="As_t92_44C11"
            /clone_lib="Ascaris suum adult male testis germinal zone
            from Alan Scott"
            /sex="Male"
            /dev_stage="Adult"
            /note="Vector: Lambda zap II; Site.1: EcoRI; Site.2: XhoI;
            Library was made from dissected testis germinal zone from
            adult male Ascaris suum collected from abattoirs.
            Constructed by Michelle Lizotte-Waniewski for Alan Scott,
            Johns Hopkins University Medical School, Baltimore, MD."
BASE COUNT  173 a      127 c      49 g      98 t
ORIGIN
Query Match      4.0%; Score 52.6; DB 10; Length 447;
Best Local Similarity 49.8%; Pred. No. 0.0092;
Matches 133; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

```

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Atan	Scott	Ascaris	sumu	CDNA	clone	As_tgz_54C05	5', mRNA	sequence.	
BM278181									
BM278181.1	GI:17971439								
EST.									
pig	roundworm.								
Ascaris	sumu								
Ascaris	sumu								
Eukaryote:	Metazoa:	Nematoda:	Chromadorea:	Ascaridida:	Ascaridoidea				
1 (bases	1 to 597)								
Blaxter	M.L.	Parkinson	J.	Whitton	C.	Daub	J.	Guiliano	D., Hall
N., Quayle	M.	and	Barrell	B.					
Edinburgh	University	Sanger	Centre	Nematode	EST	Project			
Unpublished	(2000)								
Contact:	Blaxter	ML							

```

- Plate: 54 row: C column: 05
  Seq primer: SKPL
  High quality sequence stop: 502
FEATURES
  Location/Qualifiers
    1..597
    source

```

```

/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As tgz 54C05"
/clone_1ib="Ascaris suum adult male testis germinal zone
from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI.
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Liotte-Maniowski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."

```

Query Match	3.9%	Score 52	DB 10	Length 597
Best Local Similarity	49.3%	Pred. NO.	0.015	
Matches 136	Conservative 0	Mismatches 140	Indels 0	Gaps 0

QY	431	actctcgagatatacaaacctcctgactcattcaacacgcgttgatctctgcttaccatcacca	490
Db	141	ACAGTAACCTTCACACAGCAAGCAATTTCAACAGCAACCTGAGCAACTTCATTAACAACCTTA	200
QY	491	acaactcgctcgtataaactccaaaatctacatcaacgycgcgtctcgtatgcagcaaaaaccca	550
Db	201	GCAACTCTCTGTAACAACTTCATTAACAACCTTCAGCAACCTGTCTGAGCAACCTCAATTTCA	260
QY	551	tctccaactctggtgaacatccagcgtctctataatacatcatgttcaaacctgtagcgtgttc	610
Db	261	ACTTGGAAACACCTTCACACACCTTCATTAACAACCTTAATCTAATCTCAACAACAATCTTCACCA	320
QY	611	gtgacactccacgcgtctacatctcgatgcaaatactaactcgtcttcgacaagaactgaaac	670
Db	321	ATCTCAACAATTAATCCTCAACAACCTTTACGACAGCAACTTCAACAGCAGTAACCTTCACCA	380
QY	671	aaaaagaatcaagaacctgtacgacaaacgcagtcca	706
Db	381	ACAGTAACCTTCAGCAACCTTCACAGCAACCAACTTCA	416

RESULT	8
Locus	BM278805
DEFINITION	BM278805 As_tg2-67F10-SKPL Ascaris suum adult male testis germinal zone from Alan Scott Ascaris suum cDNA clone AS-tg2-67F10 5' mRNA sequence.
ACCESSION	BM278805
VERSION	BM278805.1 GI:17972063
KEYWORDS	EST,
SOURCE	Pig Roundworm,
ORGANISM	Ascaris suum
REFERENCE	Eukaryote; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea ; Ascaridae; Ascaris. 1 (bases 1 to 500) Blaxter M.L., Parkinson,J., Whitton,C., Daub,J., Gulliano,D., Hall ,N., Quayle,M. and Barrell,B Edinburgh University/Sanger Centre Nematode EST Project Unpublished (2000)
TITLE	Contact: Blaxter ML
JOURNAL	
COMMENT	

REFERENCE	TITLE
AUTHORS	JOURNAL
	COMMENT
1 (bases 1 to 500) Blaxter/M.L., Parkinson,J., Whitton,C., Daub,J., Guiliano,D., Hall , N., Quayle,M. and Barrell,B. Edinburgh University/Sanger Centre Nematode EST Project Unpublished (2000)	
Contract: Blaxter ML Institute of Cell, Animal and Population Biology University of Edinburgh Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK Tel: +44 131 650 6760 Fax: +44 131 670 3450 Email: mark.blaxter@ed.ac.uk	
The library was prepared by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore. Sequencing was performed by the Pathogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The sequence contained a PolyA tail (trimmed)	
PCR Primers FORWARD: T3 BACKWARD: T7PL Plate: 67 row: F column: 10 Seq primer: SKPL High quality sequence stop: 500.	

FEATURES	Location/Qualifiers
source	1. .500

```

/NOTE=Vector: Lambda Zap II; Site1: EORI; Site 2: XhoI;
library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Lizotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."
BASE COUNT      193 a      58 g      105 t
ORIGIN          144 c

```

Query Match	3.9%	Score 51.8	DB 10	length 500
Best Local Similarity	50.8%	Pred. No. 0.015		
Matches 153	Conservative	0	Mismatches 142	Indels 6
				Gaps 1

QY 431 actctcggatgatcaaacctctctactcaatcaaatcgctgatctctgttccatcaaca 490
Dd 70 ACAGTAACCTTCACACAGCAATTTCAACGACCAACCTCAGAACTTCAATAACAACCTTA 129
QY 491 acaatcgctctgaataaactccaanaatctatcatcaacgycgctctgatgcacgaaccca 550
Dd 130 GCAACTCTCTTAACAACTTCAATAAACAACCTTCACCAACTCTCTGAGCAACCTTCAG 184
QY 551 tctccaatctgggttaacatccacgcttcaataaatcatgttcaaaactgtagcttgc 610
Dd 185 -CTACAACTTGAACAACCTCAACAACCTCAACCTCAACCTCAACTATTTCAACAACAATC 243
QY 611 gtgaactacacgcttaactctgatacctaacttcaactctgtcttgacaaagaactgaagc 670

Db 244 TCAGCAATCTCAACATTAACCTTCACAACACTTAGAACAAGCAACTTCAACAACAGTAACACT 303

Oy 671 aaaagaataatcaagaaacctgttagacaaccagtcacattctcgttactctgaaggcttct 730
||| ||| | | | | | | | | | | | | | | |
Db 304 TCMAAACCTTCAACGACAAACAATTCCAGCACAATTCAACAAGACGTAACTTCAACAACAGCA 363
||| ||| | | | | | | | | | | | | | | |

Oy 731 g 731
|
Db 364 G 364

RESULT 9
BM004465 511 bp mRNA linear EST 25-OCT-2001

LOCUS BM004465

DEFINITION TgSTya8f03.y1 TyVEG Partially sporulated oocyst cDNA Toxoplasma gondii cDNA clone TGESTzya86f03.y1 5' mRNA sequence.

ACCESSION BM004465

VERSION BM004465.1 GI:16439239

SOURCE EST.
Toxoplasma gondii.
Toxoplasma gondii.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
REFERENCE 1 (bases 1 to 511)
Clang,R., Cole,R., Fogarty,S., Sibley,L.D., Ajioke,J.A., White,M., Clifton,S., Pape,D., Martin,J., Wylie,T., Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,R., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvilli,R., Ronko,I., Kennedy,S., Maguire,L., Waterston,R. and Wilson,R.
Toxoplasma EST Project
Unpublished (2001)
Contact: Clifton, S.
Toxoplasma EST project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel.: 314 286 1800
Fax: 314 286 1810
Email: toxo@watson.wustl.edu
Contact David Sibley (toxest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability.
High quality sequence stop: #33.
Location/Qualifiers
1..511
/organism="Toxoplasma gondii"
/strain="TyEG (Type III)"
/db_xref="taxon:5811"
/clone="TGESTzya86f03.y1"
/clone_id="TyVEG Partially sporulated oocyst cDNA"
/dev_stage="Partially sporulated oocysts"
/note="Vector: Modified bluescript (PBSSK+); Site_1:
BamHI; Site_2: EcoRI; Polya mRNA from partially-sporulated oocysts was converted to cDNA using the template-switching PCR method (SMART cDNA, Clontech Inc.,) and sized selected on SizeStep 400 columns (Amersham Pharmacia Biotech inc.). First strand was reverse transcribed using the CDS III-oligo-dT primer and a 5' template switch primer (Smart IV primer). The product of the first strand synthesis was PCR amplified using the same primer set and the fragments were digested with SfiI. The fragments were size selected, ligated into a modified bluescript vector containing directional SfiI sites, and electroporated into DH10B or DH25 cells. Vector: SfiI sites were added to the multiple cloning region of pBluescript SK+ between the BamHI/EcoRI sites. The modified polylinker has the following sequence:
5'-GAATTCGGCGCATTTACGCC(G)n--insert--
GGCGGCCTGGCCACGAGATCC3'-where n=3-4 G nucleotides. Library Source: Michael White, Maria E. Jerome, Emily A.Johnson, Jay A. Radke, Montana State university. Clone Availability: David Sibley, Washington University"

BASE COUNT 213 a 181 c 61 g 56 t

Query Match	3.8%;	Score 50.6;	DB 10;	Length 511;
Best Local Similarity	46.9%;	Pred. No. 0.031;		
Matches 158;	Conservative	0;	Mismatches 179;	Indels 0; Gaps 0;
QY 426	caaatcattctcagatgatgatacaaatctctcgtactacaatcgaatcgctgtagatctctgtaacat	485		
Db 169	CTACTACTACAACTTCGACACCAACCAACAACTACAACTCACTCACTCACTCACTCACTCA	228		
QY 486	caccacaatcgtctcgaataactccaataacttacaatgaagggcgtctgtagcaccagaa	545		
Db 229	CACACACAAACCAACCAAACTACACCAACAGACAAACAAACCAACCAACCACTCACTCA	288		
QY 546	accgatctccaatctcggtaaatccatccacgccttctaataacataatgttccaactgagc	605		
Db 289	CAACACAAACGACAAACCAACCAACGACCACTACAGACAACTTCAGCAACCACTACGA	348		
QY 606	ttgtcgtgacacatccacgcgtacatcttgatcaaatcttaacttaacttgctgcacaaga	665		
Db 349	CTACACACACTACTCTACTCTACCACTACCAACGACAAACGACGAAACCACTCACTCA	408		
QY 666	gaacgaataaagaataccaagaacctgtacgacaacgaacgaacttcgtgtatccgtaaga	725		
Db 409	CAACACCAACGAAACCAACCAACCACTACAAAGACACAGACGACGACTTCGACAACTCA	468		
QY 726	ctctcgggtgactaccctgcagtagacaacacggtac	762		
Db 469	CTACGACACCACTCTTCGACGACGAACTCCACCACTAC	505		
RESULT 10				
LOCUS	BM278657			
DEFINITION	As.tgz.65B06.SKPL Ascaris suum adult male testis germinal zone from			
ACCESSION	BM278657			
VERSION	BM278657.1			
KEYWORDS	EST.			
SOURCE	plg roundworm.			
ORGANISM	Ascaris suum			
REFERENCE	Eukaryotes: Metazoa: Nematoda; Chromadorea; Ascaridida; Ascaridoldea			
AUTHORS	; Ascarididae: Ascaris.			
TITLE	1 (bases 1 to 569)			
JOURNAL	Blaxter, M.L., Parkinson, J., Whitton, C., Daub, J., Guiliano, D., Hall			
COMMENT	, N., Quayle, M. and Barrell, B. Edinburgh University/Sanger Centre Nematode EST Project Unpublished (2000) Contact: Blaxter ML Institute of Cell, Animal and Population Biology University of Edinburgh Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JY, UK. Tel: +44 131 650 6760 Fax: +44 131 670 5450 Email: mark.blaxter@ed.ac.uk The library was prepared by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore. Sequencing was performed by the Pathogen Sequencing Unit, Sanger Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The sequence contained a Polya tail (trimmed) PCR Primers FORWARD: T3 BACKWARD: T7PL Plate: 65 row: E column: 06 Seq primer: SKPL High quality sequence stop: 518. Location/Qualifiers 1..569 /organism="Ascaris suum" /db_xref="taxon:6253" /clone="As.tgz.65B06" /clone_lib="Ascaris suum adult male testis germinal zone from Alan Scott"			


```
/sex="Male"
/dev-stage="Adult"
/!note="Vector: Lambda Zap II; Site.1: EcoRI; Site.2: XhoI;
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Lizotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."
BASE COUNT      213 a      164 c      69 g      123 t
ORIGIN

Query Match      3.8%; Score 50.4; DB 10; Length 569;
Best Local Similarity 48.9%; Pred. No. 0.037;
Matches 135; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 431 actctcagatgatacaatctctgataataatctgctgatactctgatactacac 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 119 ACAGTACTTGAACCAACCACTTTCACAGCAACCTCAGCAACTTCAATTAACCACTTGA 178
QY 491 acaatcgtctgatacaatcaaatctacatacaagcgctctgatacaggaacgca 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 179 GCAACTCTTGAAACCACTTCAATTAACCACTCAGCAACTCTGAGCAACCTCAGCTACA 238
QY 551 tctcaatctggttaacatccacgctcttaataacatcatgtaaacctgacggtgttc 610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 239 ACTTGAAACCACTTCAACCACTTCAATTAACCACTCAGCAACTTCAATTAACCACTCAGCA 298
QY 611 gtgacatcactcagctacatctgataataatctcaatctgctgatacagaactgaag 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 299 ATCTCAACCAATTAACCTTCAACCACTTGAAGCAAGCAACTTCAACCACTTCAATTAACCTTCA 358
QY 671 aaaaagaatcaagaagcgtctgatacagcagctca 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 359 ACAGTACTTGAACCACTTCAACCACTTCAACCACTTCA 394

RESULT 11
BM181884      641 bp      mRNA      linear      EST 11-DEC-2001
LOCUS      fv51b11.y1 Sugano SJD adult male Dario rerio cDNA clone 5412044 5'
DEFINITION      similar to contains element TAK1 repetitive element ;, mRNA
sequence.
ACCESSION      BM181884
VERSION      BM181884.1 GI:17512842
KEYWORDS      EST
SOURCE      zebrafish.
ORGANISM      Dario rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE      1 (bases 1 to 641)
AUTHORS      Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood
,K., Stepoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
TITLE      Washu zebrafish EST Project 1998
JOURNAL      Unpublished (1998)
COMMENT      Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbratfish@watson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clone
distribution information can be found through the I.M.A.G.E.
Consortium/LNL. send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 527.
FEATURES
SOURCE      1..641
Location/Qualifiers
```

```
/organism="Dario rerio"
/db_xref="taxon:7955"
/clone="5412044"
/clone_1ib="Sugano SJD adult male"
/sex="male"
/issue_type="whole body"
/dev-stage="adult"
/!note="Vector: pME185-FL3; Site.1: DraIII (CAACCAATG);
Site.2: DraIII (CACTGCTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGAGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TTGAGCCCTTCTG], digested and cloned into distinct DraIII
sites of the pME185-FL3 vector (5' site CACTGCTG, 3' site
CAACCAATG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed and donated by Dr. Sumio
Sugano (University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTGTGCTCTTAAGCTGGC and 3' end primer
CGACCTGCAGCTCGAGCACA."
BASE COUNT      218 a      234 c      37 g      152 t
ORIGIN

Query Match      3.8%; Score 50; DB 10; Length 641;
Best Local Similarity 46.3%; Pred. No. 0.049;
Matches 201; Conservative 0; Mismatches 230; Indels 3; Gaps 1;

QY 229 tacaactctatgacgaagaactctcaactctctctgatacgtatccgaatacttc 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 175 TACATCAACTTTCACCAACCAACCACTTCTCCACCAACCAACCACTTCTCCACTCATC 234
QY 289 aactcactctctgacaaatgataacacatcatcaactgactggaacaa---ttct 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 235 AACCTCAACCAACCAACCAACCACTTCTCCAGCTACATCAACTTCAACCAACCACTTCTCC 294
QY 346 ggttggaagatctctgatacagtgataatctgatactgatacagcagcagcagga 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 295 AGCTTACCACTTTCACCAACCAACCACTTCTCCAGCTACATCAACTTCTCCACCAAC 354
QY 406 atcaaacagcgtgtgtatcaataactctcagatgatcaacatctctgatacatcaat 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 AACCACTTTCACCAACCAACCACTTCTCCACCAACCAACCACTTTCACCACTCATC 414
QY 466 cgtctgatacttgatcaccacaaacatcgtctgaataactccaaatctacatcaac 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 415 TTCTCCACCAACCAACCACTTCTCCAGCTACATCAACTTCTCCACCAACCAACCACTTCTCC 474
QY 526 ggcgcgtctgatacaggaacacgcatctcaatctggtgataacatccagcttctaatac 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 475 AGCTTACATCACTTCAACCAACCAACCACTTCTCCAGCTACATCAACTTCTCCACCAACCAAC 534
QY 586 atcatgtcaaatgagcgtgtgtcgtgatacactcagcagctacatctgatacaatacttc 645
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 535 AACTTCTCCAGCTACATCAACTTCAACCAACCAACCAACCACTTCTCCAGCTACATCAACTTTC 594
QY 646 aatctgttgacaa 659
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 595 AACCAACCAACCAAA 608

RESULT 12
A1389106      645 bp      mRNA      linear      EST 23-APR-2001
LOCUS      GH20192.5prime GH Drosophila melanogaster head port2 Drosophila
DEFINITION      melanogaster cDNA clone GH20192 5, mRNA sequence.
ACCESSION      A1389106
VERSION      A1389106.2 GI:13758704
KEYWORDS      EST.
SOURCE      fruit fly.
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
```

REFERENCE 1 (bases 1 to 645)
 AUTHORS Harvey, D., Brokstein, P., Hong, L., Evans-Holm, M., Su, C., Tsang, G., Lewis, S., and Rubin, G. M.
 TITLE BDGP/HMMI Drosophila EST Project
 JOURNAL Unpublished (2001)
 COMMENT On Jan 28, 1999 this sequence version replaced gi:4203117.
 Other-ESTs: GH20192.3prime
 Contact: Stapleton, M.
 BDGP

Lawrence Berkeley National Lab
 One Cyclotron Rd, Berkeley, CA 94720, USA
 Fax: 510 486 6798
 Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu
 Plate: GH.201 row: H column: 8
 High quality sequence stop: 627
 POLY-A-NO.

FEATURES

source 1..645
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="GH20192"
 /clone_1lb="GH Drosophila melanogaster head port2"
 /sex="male and female"
 /dev_stage="adult"
 /lab_host="DH5 - alpha"
 /note="Organ: head; Vector: PORT2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into port2. Plasmid cDNA library."
 BASE COUNT 239 a 232 c 94 g 80 t
 ORIGIN

Query Match 3.7%; Score 49.6; DB 9; Length 645;
 Best Local Similarity 47.0%; Pred. No. 0.062;
 Matches 154; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 482 ccataccaacaatcgtcgaataatccaataatcacaacgcgcgtcgtacgacc 541
 DB 209 CCACCACTGTCTAGGGAACACGACCAACACGAGTTCCTTACCAACCAACCA 268
 QY 542 agaacgcgactccaatcgtgtaacatccaagcttctaataatattgtaactcg 601
 DB 269 ACGTGTCCCACTACCACTGACCACTACCACTACCACTACCACTACCACTAC 328
 QY 602 acggttcgtgacacccgcgtacacacacacacacacacacacacacacacac 661
 DB 329 ACTACGAC 388
 QY 662 aactgaacgaagaagaatacaagaactgtacgaacacacacacacacacacac 721
 DB 389 ACCAC 448
 QY 722 aagaacttcgtggtactactcgtgaagtaacgaacacacacacacacacacac 781
 DB 449 ACACGAC 508
 QY 782 atccgaacaataacgttgcgtacaa 809
 DB 509 AC 536

RESULT 13
 CDS02156 827 bp DNA linear GSS 12-MAY-2000
 LOCUS Tetradon nigroviridis genome survey sequence pUC-ori end of clone
 DEFINITION 224F10 of library G from Tetradon nigroviridis, genomic survey
 sequence.
 ACCESSION AL176451
 VERSION AL176451.1 GI:7814508
 KEYWORDS GSS: genome survey sequence.
 SOURCE Tetradon nigroviridis.

ORGANISM

Tetradon nigroviridis
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes;
 Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 827)
 AUTHORS Roest-Crolius, H., Jallion, O., Dasilva, C., Fizes, C., Fisher, C., Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
 TITLE Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 827)
 AUTHORS Roest-Crolius, H., Jallion, O., Dasilva, C., Bonneau, L., Fisher, C., Bernot, A., Fizes, C., Wincker, P., Brothier, P., Quetier, F., Saurin, W. and Weissenbach, J.
 COMMENT Human gene number estimate provided by genome wide analysis using Tetradon nigroviridis DNA sequence

Unpublished
 3 (bases 1 to 827)

REFERENCE 3 (bases 1 to 827)
 AUTHORS Direct Submission
 COMMENT Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
 This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetradon>.

FEATURES

source 1..827
 /organism="Tetradon nigroviridis"
 /db_xref="taxon:99883"
 /clone="224F10"
 /clone_1lb="G"
 /note="Genoscope sequence ID : C0AG224DC05SP1-end ; pUC-ori"
 BASE COUNT 368 a 203 c 106 g 121 t 29 others
 ORIGIN

Query Match 3.7%; Score 49.2; DB 12; Length 827;
 Best Local Similarity 43.8%; Pred. No. 0.086;
 Matches 169; Conservative 9; Mismatches 208; Indels 0; Gaps 0;

QY 426 caaactctcagatgacacatctctgactacataatcgtcgtatctgtaacat 485
 DB 166 CACGAC 225
 QY 486 caccacaatcgtctgataactccaatctacataacacgcgtctgacgcagaa 545
 DB 226 CAC 285
 QY 546 accgatccaatcgtgtaacatccacgcttccaataacatcgttcaactgcg 605
 DB 286 TAACACACACATTAACAGCSTATTAACACATTAACAGCSTATTAACAC 345
 QY 606 ttgtcgtgaactcaccgtacatctggaatcaactcaatctgttcgaagaact 665
 DB 346 CAACATTAACACACATTAACAGCSTATTAACACATTAACAGCSTATTA 405
 QY 666 gaacgaagaagaatacaagaacacgttcgaacacacacacacacacacac 725
 DB 406 CAATTAACAGTGTATTAACATTAACACATTAACACATTAACACATTA 465
 QY 726 ctctcgtggtgactactcgtcgtacgaacacacacacacacacacacacac 785
 DB 466 TGACAGCGATGRCACACACACACACACACACACACACACACACACAC 525
 QY 786 gaacaatacgttgcgtcaacatg 811
 DB 526 CAACATTAACAGTGTATTAACACATTA 551

RESULT 14

